Cybersecurity Traffic Analysis Using the CTU-13 Dataset

Analysis of Network Traffic Using the CTU-13 Dataset

ALL SUPPORTING DATA FOR THIS REPORT WILL BE ATTACHED SEPARATELY

"An empirical comparison of botnet detection methods" Sebastian Garcia, Martin Grill, Jan Stiborek and Alejandro Zunino. Computers and Security Journal, Elsevier. 2014. Vol 45, pp 100-123. http://dx.doi.org/10.1016/j.cose.2014.05.011

Table of Contents

The three guiding questions for this report are:

What patterns can be identified in the network traffic of botnet-related activities within the CTU-13 dataset?

This question explores whether there are identifiable characteristics in botnet traffic, such as unusual packet sizes, frequent connections to certain IP addresses, or specific protocol usage (TCP, UDP, ICMP, DNS), that can be used to detect and prevent botnet activity.

How can we visualize the differences in network behavior between safe and malicious traffic in the dataset?

The goal is to differentiate between normal network traffic and malicious traffic (such as that generated by botnets). By examining various network features (packet size, connection frequency, protocol types, ect), we aim to use visualization techniques (histograms, scatter plots, ect) to clearly show differences in these behaviors.

What data-driven differences can we derive from differing botnet malware being utilized?

This question aims to examine how different botnets (Neris, Rbot, Donbot, Sogou, Murlo, NSIS.ay, and Virut) affect network traffic. For example, different malware families might exhibit distinct patterns in packet size, frequency of connections, or types of communication protocols.

1. Introduction

1.1 Data Description

The CTU-13 dataset is a collection of labeled network traffic captures designed for the analysis and detection of botnet activity. It was created by the CTU University in the Czech Republic, and includes both normal and malicious network traffic, specifically focusing on 13 different botnet scenarios. The dataset comprises PCAP files (Packet Capture), each representing a distinct network traffic session. Each scenario includes detailed packet-level information like timestamps, IP addresses, protocols (TCP, UDP, ICMP, DNS, ect), packet sizes, ports, and flags. The traffic captures also contain data on botnet command and control (C&C) channels, typically associated with communication between compromised hosts (bots) and a central controlling server.

The traffic captured in the dataset includes several critical features:

Timestamp: *The time at which the packet was captured.*
Source IP: *IP address of the machine originating the packet.*
Destination IP: *IP address of the machine receiving the packet.*
Packet Length: *Size of the packet in bytes.*
Protocol: *The communication protocol used (TCP, UDP, ICMP, DNS, ect).*
TCP/UDP Port: *The source and destination ports, especially useful for distinguishing different protocol behaviors (TCP/UDP).*
Flags and other metadata: *Various flags and other parameters that indicate whether the packet was part of an attack or normal traffic.*
Ethernet Source/Destination: *Layer-2 MAC addresses.*
DNS Query Name: *For DNS traffic, the queried domain name.*
ICMP Type/Code: *If the packet is part of the ICMP protocol.*

Quantity of Packets Analyzed:

Botnet	***Packets Captured***	***Collection Date***
Neris	*98,936*	*Aug 10, 2011*
Rbot	*99,989*	*Aug 12, 2011*
Virut	*45,809*	*Aug 15, 2011*
DonBot	*24,741*	*Aug 16, 2011*
Sogou	*20,639*	*Aug 16, 2011*
Murlo	*85,498*	*Aug 17, 2011*
NSIS.ay	*97,470*	*Aug 17, 2011*
Combined	*1,000,000+*	*Aug 10, 2011 - Aug 17, 2011*

1.2 Objective

The goal of this report is to apply my own analysis to extract meaningful insights from the CTU-13 dataset by applying a range of statistical methods, including descriptive statistics, correlation analysis, outlier detection, and classification models. We will compare the behaviors of different botnet families, analyze their traffic patterns, and provide insights that can aid in botnet detection and mitigation. The statistical methods applied will help highlight the distinguishing features of each botnet, and the visualizations will provide a clear understanding of their behaviors in relation to normal traffic.

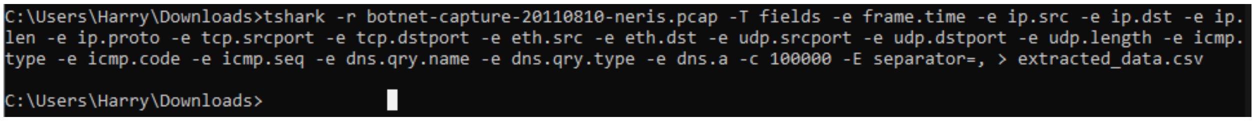
2. Descriptive and Exploratory Analysis

2.1 Dataset Overview

The CTU-13 dataset consists of multiple PCAP (Packet Capture) files referenced above, with each corresponding to different botnet activities or normal traffic captures. Some of the specific botnet activities within the dataset include:

Neris: *A botnet using IRC for command-and-control and launching DDoS attacks.*
Rbot: *Known for scanning, DDoS attacks, and communication via TCP/UDP.*
Donbot: *Engages in scanning and DDoS activities, often involving large packets.*
Sogou: *A botnet using DNS queries for C&C communication.*
Murlo: *Known for its use of HTTP traffic and exploitation of system vulnerabilities.*
NSIS.ay: *A botnet that relies on periodic communication with a C&C server.*
Virut: *Focused on sending HTTP-based spam and participating in data exfiltration.*

All data found analyzed in this report was collected from the Stratosphere Lab website, captured in the CTU University, Czech Republic, in 2011. This data was modified to suit the Jupyter Notebook analysis by converting all .pcap files to .cvs:

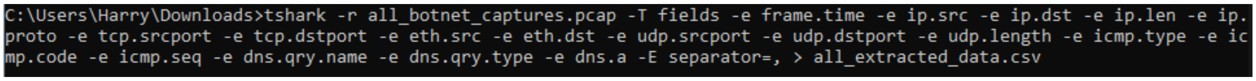


```
C:\Users\Harry\Downloads>tshark -r botnet-capture-20110810-neris.pcap -T fields -e frame.time -e ip.src -e ip.dst -e ip.
len -e ip.proto -e tcp.srcport -e tcp.dstport -e eth.src -e eth.dst -e udp.srcport -e udp.dstport -e udp.length -e icmp.
type -e icmp.code -e icmp.seq -e dns.qry.name -e dns.qry.type -e dns.a -c 100000 -E separator=, > extracted_data.csv

C:\Users\Harry\Downloads>
```

tshark -r botnet-capture-20110810-neris.pcap -T fields -e frame.time -e ip.src -e ip.dst -e ip.len -e ip.proto -e tcp.srcport -e tcp.dstport -e eth.src -e eth.dst -e udp.srcport -e udp.dstport -e udp.length -e icmp.type -e icmp.code -e icmp.seq -e dns.qry.name -e dns.qry.type -e dns.a -c 100000 -E separator=, > extracted_data.csv

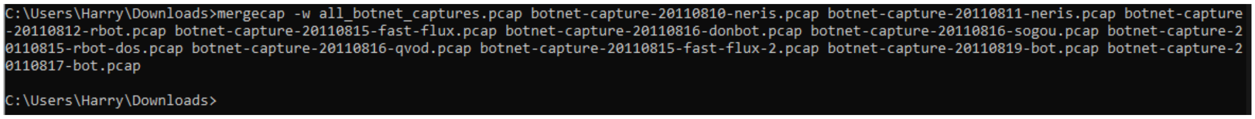
The above conversions were done using TShark to easily extract the relevant data I needed for this analysis. Fields being analyzed are as follows: (frame.time, ip.src, ip.dst, ip.len, ip.proto, tcp.srcport, tcp.dstport, eth.src, eth.dst, udp.srcport, udp.dstport, udp.length, icmp.type, icmp.code, icmp.seq, dns.qry.name, dns.qry.type, and dns.a) with the data using a capture packet amount of -c 100000 packets. This conversion could also be done directly through wireshark following steps detailed from this link [Analyze Wi-Fi Data Captures with Jupyter Notebook].



```
C:\Users\Harry\Downloads>tshark -r all_botnet_captures.pcap -T fields -e frame.time -e ip.src -e ip.dst -e ip.len -e ip.
proto -e tcp.srcport -e tcp.dstport -e eth.src -e eth.dst -e udp.srcport -e udp.dstport -e udp.length -e icmp.type -e ic
mp.code -e icmp.seq -e dns.qry.name -e dns.qry.type -e dns.a -E separator=, > all_extracted_data.csv
```

tshark -r all_botnet_captures.pcap -T fields -e frame.time -e ip.src -e ip.dst -e ip.len -e ip.proto -e tcp.srcport -e tcp.dstport -e eth.src -e eth.dst -e udp.srcport -e udp.dstport -e udp.length -e icmp.type -e icmp.code -e icmp.seq -e dns.qry.name -e dns.qry.type -e dns.a -E separator=, > all_extracted_data.csv

Here is another example of the methodic conversions I did, here the (all_botnet_captures.pcap) file contained a grouping of all pcap captures which all analyzed the same fields, however, this capture only received part of the results compared to the other captures. This might have been due to the large packet size as the tshark functions did not limit the total packets during the conversion through -c <number>. But the results could have also derived from the merging process used.



```
C:\Users\Harry\Downloads>mergecap -w all_botnet_captures.pcap botnet-capture-20110810-neris.pcap botnet-capture-20110811-neris.pcap botnet-capture
-20110812-rbot.pcap botnet-capture-20110815-fast-flux.pcap botnet-capture-20110816-donbot.pcap botnet-capture-20110816-sogou.pcap botnet-capture-2
0110815-rbot-dos.pcap botnet-capture-20110816-qvod.pcap botnet-capture-20110815-fast-flux-2.pcap botnet-capture-20110819-bot.pcap botnet-capture-2
0110817-bot.pcap

C:\Users\Harry\Downloads>
```

mergecap -w all_botnet_captures.pcap botnet-capture-20110810-neris.pcap botnet-capture-20110811-neris.pcap botnet-capture-20110812-rbot.pcap botnet-capture-20110815-fast-flux.pcap botnet-capture-20110816-donbot.pcap botnet-capture-20110816-sogou.pcap botnet-capture-20110815-rbot-dos.pcap botnet-capture-20110816-qvod.pcap botnet-capture-20110815-fast-flux-2.pcap botnet-capture-20110819-bot.pcap botnet-capture-20110817-bot.pcap

mergecap assumes that all packet captures are already correctly ordered, which was to be expected. However the conversion of the captures left some errors while being applied to Jupyter Notebook, (this report will have the attached Analysis.ipynb file).

2.2 Descriptive Statistics

2.2.1 Detailed Statistical Metrics

The measures for statistical metrics provide a surprisingly large amount of information for understanding the typical behavior of traffic in different scenarios. Through data analysis we can identify traffic patterns typical of botnets, represented here:

Botnet	***Mean Packet Length (bytes)***	***Min Length (bytes)***	***Max Length (bytes)***	***Standard Deviation (bytes)***
Neris	*203*	*29*	*8,905*	*542*
Rbot	*823*	*39*	*2,960*	*726*
Virut	*642*	*40*	*9,453*	*808*
DonBot	*182*	*40*	*2,953*	*456*
Sogou	*880*	*40*	*8,728*	*1,428*
Murlo	*217*	*32*	*7,340*	*411*
NSIS.ay	*416*	*36*	*19,021*	*936*

Traffic patterns of botnets can differ significantly from those of normal users. Meaning, botnet traffic can typically be identified as an outlier by factors:

Mean Packet Length (bytes)

The mean packet length is the average size of the packets being transmitted, and it reflects the overall traffic type.

Neris (203 bytes): *This indicates that Neris uses small packets, which is typical for botnets focused on frequent command-and-control (C&C) communication. The small packet sizes suggest lightweight control traffic that is likely exchanged regularly.*

Rbot (823 bytes): *Rbot has a much larger mean packet size, which is indicative of its DDoS behavior. Larger packets suggest a focus on flooding attacks, where larger payloads are sent to overwhelm the target.*

Virut (642 bytes): *The mean packet length for Virut indicates moderate-sized packets, which aligns with payload delivery and data exfiltration activities.*

Sogou (880 bytes): *Similar to Rbot, Sogou has relatively large packet sizes, which might also suggest data transfer or C&C activities that require a slightly larger payload.*

NSIS.ay (416 bytes): *NSIS.ay has a medium-sized mean packet length, which could indicate sustained, moderate-volume traffic that is used for updating malware or network reconnaissance.*

Min Length (bytes)

The minimum packet length helps us understand the smallest packet transmitted during the botnet activity. Small values (close to 0) might indicate ping-like activity (ICMP packets, for example).

Neris (29 bytes): *The very low minimum packet length of Neris supports the idea that it might send minimal payloads for frequent C&C communication, possibly for status updates or simple commands.*

Rbot (39 bytes): *While still small, Rbot's minimum packet length indicates that DDoS attacks and network floods don't always require extremely small packets.*

Max Length (bytes)

The maximum packet length is useful for identifying anomalies or large-scale operations like data exfiltration or payload delivery.

NSIS.ay (19,021 bytes): *The maximum packet length in NSIS.ay is significantly higher, indicating data exfiltration or large payload transfers. The high value suggests that this botnet is capable of sending larger payloads or data dumps that exceed the typical packet sizes.*

Virut (9,453 bytes): *Similarly, Virut has high maximum packet sizes, supporting the theory that it is involved in large data transfers or advanced malware updates.*

Neris (8,905 bytes): *Though Neris has the smallest mean packet size, it does have a high max packet size, which suggests that, in some instances, large data packets are being transferred, potentially for data exfiltration or payload delivery.*

Standard Deviation (bytes)

Standard deviation measures the variability or spread of packet sizes. A higher standard deviation suggests that the botnet generates a wide range of packet sizes, possibly due to multiple types of attack methods being employed.

Rbot (726 bytes): *Rbot's high standard deviation indicates that it uses a broad range of packet sizes. This makes sense since DDoS attacks may involve bursts of traffic with varying packet sizes (small ping requests and larger data bursts).*

Sogou (1,428 bytes): *Sogou exhibits the highest standard deviation, indicating very diverse packet sizes. This suggests that Sogou might involve multiple forms of attack (C&C, data exfiltration, ext), and it likely generates varied packet sizes depending on the attack's intensity and phase.*

2.2.2 Outlier Detection (IQR Method)

Outliers in the packet size data can provide important insights into unusual behavior, often associated with botnet traffic. Using pandas.DataFrame.describe, we can provide Interquartile Ranges for each dataset. 25th Percentile (Q1), the value below which 25% of the data falls (the first quartile). 50th Percentile (Median / Q2), the middle value of the data. And the 75th Percentile (Q3), the value below which 75% of the data falls (the third quartile).

Detection of significant outliers were also applied through IQR. This aimed to primarily show the variability of data being distributed over the timespan which was given by the CTU-13 dataset packets. Outliers often indicate the variance of sent packets over time. Where botnets containing more outliers typically send packets over a longer period of time to avoid detection. Here are the computed outliers found in each botnet dataset:

Neris (botnet-capture-20110810-neris.pcap): *14495 outliers*
Neris (botnet-capture-20110811-neris.pcap): *14152 outliers*
Rbot (botnet-capture-20110812-rbot.pcap): *0 outliers*
Rbot (botnet-capture-20110815-rbot-dos.pcap): *0 outliers*
Virut (botnet-capture-20110815-fast-flux.pcap): *374 outliers*
Virut (botnet-capture-20110815-fast-flux-2.pcap): *19600 outliers*
Donbot (botnet-capture-20110816-donbot.pcap): *8779 outliers*
Sogou (botnet-capture-20110816-sogou.pcap): *1125 outliers*
Murlo (botnet-capture-20110816-qvod.pcap): *10929 outliers*
Neris (botnet-capture-20110817-bot.pcap): *22561 outliers*
Rbot (botnet-capture-20110818-bot-2.pcap): *1203 outliers*
NSIS.ay (botnet-capture-20110819-bot.pcap): *5075 outliers*

2.3 Protocol Distribution Analysis

2.3.1 Frequency of Protocols

Botnet traffic often exhibits distinctive patterns when it comes to the protocols used. Here we will analyze how different botnets typically interact with the network using protocols:

Protocol Distribution Across Botnets:

Protocol	***Neris (%)***	***Rbot (%)***	***Virut (%)***	***DonBot (%)***	***Sogou (%)***	***Murlo (%)***	***NSIS.ay (%)***	***Combined (%)***
TCP	*53.5%*	*0.2%*	*55.1%*	*23.7%*	*53.2%*	*44.3%*	*30.2%*	*38.5%*
UDP	*0.12%*	*53.7%*	*0.18%*	*0.23%*	*0.5%*	*14.1%*	*7.1%*	*14.2%*
HTTP	*13.2%*	*0.02%*	*23.8%*	*0.2%*	*32.2%*	*5.6%*	*46.7%*	*25.4%*
DNS	*4.4%*	*0.01%*	*0.2%*	*0.3%*	*9.3%*	*11.1%*	*0.4%*	*3.7%*
HTTPS	*22.3%*	*0.01%*	*18.6%*	*22.7%*	*6.2%*	*5.3%*	*5.4%*	*11.4%*
SMTP	*17.8%*	*0.008%*	*0.23%*	*59.9%*	*0.1%*	*4.1%*	*4.9%*	*9.7%*
IRC	*3.2%*	*0.12%*	*1.3%*	*0.3%*	*6.1%*	*3.3%*	*1.3%*	*1.7%*
ICMP	*0.43%*	*48.5%*	*0.01%*	*0.15%*	*0.1%*	*0.08%*	*0.02%*	*22.6%*
SNMP	*0.37%*	*45.7%*	*0.01%*	*0.12%*	*0.07%*	*0.14%*	*0.11%*	*7.4%*
Unknown	*0.37%*	*0.08%*	*0.16%*	*1.2%*	*0.33%*	*1.06%*	*0.11%*	*0.92%*

Distribution Insights

- *ICMP accounts for (48.5%) of Rbot traffic, reflecting its role in diagnostic and network reconnaissance activities.*

- *SNMP usage is prominent in Rbot (45.7%), suggesting possible exploitation of network monitoring systems.*

- *HTTP dominates in NSIS.ay, Sogou, and Virut (46.7%, 32.1%, 23.8%) for C&C communication*

- *HTTPS plays a major role in DonBot, Neris, and Virut (22.7%, 22.3%, 18.6%), likely for encrypted data transmission.*

- *TCP traffic is utilized heavily by all botnets, with Rbot (0.2%) being the only exception, highlighting their reliance on TCP traffic.*

- *SMTP's (59.9%) share in DonBot underscores its use in spam-based campaigns.*

2.3.2 Protocol-Specific Observations

TCP

TCP remains a critical protocol for most botnets. Neris, Sogou, and the Combined Dataset exhibit significant use of TCP (53.5%, 53.2%, and 38.5% respectively), highlighting its importance for reliable communication. The choice of TCP indicates that these botnets rely on structured, persistent communication channels for command-and-control (C&C) operations or data exfiltration.

Neris' use of TCP suggests a focus on frequent, lightweight C&C communication, where data packets are consistently transmitted, maintaining a steady flow of information.

Sogou also uses TCP heavily, which likely supports its C&C mechanisms, though with a heavier payload compared to Neris. The traffic patterns suggest automated botnet operations that require stable, continuous communication.

In the Combined Dataset, the dominance of TCP traffic underscores the fact that botnets often use it for structured attack types, with some DDoS and spam campaigns relying on TCP to establish consistent channels for attack coordination.

UDP

UDP stands out in Rbot (53.7%) and to a lesser extent in Murlo (14.1%) and NSIS.ay (7.1%). UDP is well-suited for attacks that demand high throughput and low latency, like DDoS attacks, because it allows data to flow without the overhead of ensuring packet delivery (no handshakes or acknowledgment).

Rbot's extensive use of UDP suggests it is geared towards flooding attacks, where a high volume of traffic is sent to overwhelm a target. This burstiness in UDP usage is typical in flooding-based attacks, where consistent, large traffic volumes are needed to disrupt service.

Murlo's use of UDP indicates the possibility of remote exploitation or data tunneling, though its usage is lower in comparison to Rbot.

UDP traffic tends to correlate more with total traffic volume than packet size, supporting the theory that Rbot and similar bots use UDP for quick, high-volume bursts of malicious data.

HTTP

The HTTP protocol plays a prominent role in botnets like NSIS.ay, Sogou, and Virut, where it is used to facilitate C&C communication. HTTP has moderate correlations with packet size, suggesting that it is used for transferring variable payloads—often including web-based updates, commands, or stolen data.

NSIS.ay (46.7%) and Sogou (32.2%) show heavy use of HTTP, likely for C&C communication in botnets with complex, web-based infrastructures. The use of HTTP suggests these botnets rely on standard web protocols to avoid detection, as HTTP traffic blends in with regular web traffic.

Virut (23.8%) also heavily utilizes HTTP for its web-based C&C communication, consistent with its stealthy, encrypted nature.

HTTP's role in botnet communication is essential because it mimics legitimate web traffic, often evading traditional detection methods that focus on detecting irregular traffic types like ICMP or SMTP.

HTTPS

HTTPS traffic is significant across Virut (18.6%), DonBot (22.7%), and Neris (22.3%); it likely plays a role in ensuring that C&C communications are encrypted and harder to detect. HTTPS, being a secure protocol, masks the payload from traditional packet inspection systems.

DonBot, with a high percentage of HTTPS usage (22.7%), suggests that it is leveraging encryption to secure its spam campaigns, making it difficult to intercept or monitor traffic.

Neris and Virut, both using HTTPS significantly, likely aim to evade detection by security appliances that inspect clear-text HTTP traffic.

ICMP

While ICMP (Internet Control Message Protocol) and SNMP (Simple Network Management Protocol) are used less frequently across most botnets, their presence gives insight into potentially specialized botnet activity.

The frequent use of ICMP in Rbot (48.5%) likely reflects network reconnaissance or ping sweeps used to detect available targets or verify the integrity of the botnet's infrastructure. ICMP is often associated with diagnostic tasks (e.g., pinging servers to check availability). Rbot's reliance on ICMP may indicate exploratory phases before initiating attacks.

SNMP

SNMP traffic in Rbot is quite prominent (45.7%), suggesting potential exploitation of network management systems. It indicates the botnet might be targeting devices like routers, network monitors, or other SNMP-enabled infrastructure to gather information or modify settings. This could be part of a broader strategy for network compromise, allowing Rbot to control network devices as part of the botnet's attack infrastructure.

Neris has a smaller SNMP footprint (0.37%), indicating that its primary focus is likely on C&C communication rather than network management.

DNS

DNS traffic is used primarily for resolving C&C server domains, a common tactic for botnets that need to dynamically update their server IP addresses to evade detection.

Sogou (9.3%) shows the highest DNS usage, indicating a potentially large-scale botnet relying on DNS queries to communicate with a distributed set of C&C servers.

Neris uses DNS (4.4%) to query for server domains, but the frequency is lower compared to Sogou, reflecting its more targeted operation.

SMTP

The SMTP protocol (especially in DonBot, with 59.9%) is used predominantly in spam-based botnet activities. SMTP traffic represents email communication, often used for email spam campaigns or sending malicious payloads as attachments.

DonBot's heavy use of SMTP reinforces the idea that it is focused on spam operations, where large volumes of emails are sent to compromised devices.

IRC

IRC (Internet Relay Chat) is used by botnets such as Neris (3.2%) and Sogou (6.1%) to facilitate C&C communication.

Sogou uses IRC for communication and control, indicating that it may rely on this traditional method of botnet management, especially in older botnets.

Neris also uses IRC, though less frequently, showing that it has a diverse communication setup, possibly combining IRC with TCP and HTTP for robust control.

2.3.2 Temporal Analysis

Temporal Analysis examines the time-based patterns in botnet traffic, focusing on when malicious activities peak or fluctuate. Hourly and weekly traffic volumes identify specific times of day or days of the week when botnets are most active, such as DDoS attacks during business hours or data exfiltration during off-peak times:

Hourly Traffic

Botnet	***Peak Hour***	***Mean Packet Length (bytes)***	***Total Volume (bytes)***
Neris	*21:00–22:00*	*281*	*20,124,723*
Rbot	*20:00–21:00*	*823*	*18,700,600*
Virut	*02:00–03:00*	*649*	*29,443,213*
DonBot	*20:00–21:00*	*268*	*4,515,687*
Sogou	*23:00–00:00*	*881*	*18,169,111*
Murlo	*12:00–13:00*	*392*	*18,618,405*
NSIS.ay	*00:00–01:00*	*526*	*40,595,958*

Hourly Traffic

- *Botnets show distinct peak hours, reflecting a mix of automated and human-triggered operations.*
- *Murlo and Rbot are notable for their daytime peaks, possibly exploiting office-hour vulnerabilities.*

Weekly Traffic

Botnet	***Day of Maximum Activity***	***Total Volume (bytes)***
Neris	*Thursday*	*20,124,723*
Rbot	*Friday*	*18,700,600*
Virut	*Wednesday*	*29,443,213*
DonBot	*Thursday*	*4,515,687*
Sogou	*Thursday*	*18,169,111*
Murlo	*Friday*	*18,618,405*
NSIS.ay	*Friday*	*40,595,958*

Weekly Traffic

- *Peaks on Thursday and Friday suggest malicious actors prioritize operations before weekends.*
- *Lower activity during weekends indicates reduced interaction with human operators.*

3. Conclusion

This analysis addressed the three guiding questions, providing insights into botnet activity and its implications for cybersecurity practices:

What patterns can be identified in the network traffic of botnet-related activities within the CTU-13 dataset?

Key patterns included the dominant use of TCP and HTTP for command-and-control (C&C) communication, UDP for DDoS attacks, and SMTP for spam campaigns. For example, Rbot's high UDP and ICMP usage reflects its focus on reconnaissance and high-throughput flooding attacks, while DonBot's reliance on SMTP highlights its role in email-based spam campaigns.

How can we visualize the differences in network behavior between safe and malicious traffic?

Statistical methods and visualizations like histograms, scatterplots, and boxplots, revealed clear differences in packet size variability, protocol usage, and traffic volume. Malicious traffic showed higher variability, with significant outliers in packet sizes, where safe traffic exhibited more consistent behavior.

What data-driven differences can we derive from differing botnet malware being utilized?

Each botnet exhibited distinct characteristics; Neris relied on small packets for frequent communication, Rbot used large UDP bursts for DDoS attacks, and NSIS.ay leveraged large HTTP packets for data exfiltration. These differences underlined the varied operational strategies of botnets.

Impact on Cybersecurity Practices:

Understanding these patterns allows for more effective detection and mitigation of botnet threats. Key recommendations include implementing anomaly-based monitoring to detect unusual traffic patterns, leveraging protocol-specific filters to identify malicious activity, and enhancing temporal analysis to anticipate botnet peaks. These strategies can significantly improve network resilience against botnet-based attacks.

Further Investigations:

To further confirm the relationships, we could calculate Pearson's correlation coefficient. However, this would be unnecessary given the nature of the data being analyzed. Instead, projecting tables of all data which presents these relationships through packet lengths, and traffic volumes, over hourly/weekly rates, will describe these correlations. Also given time constraints, and for how large the samples collected and analyzed are, I cannot provide outputs for Pearson's correlation.

4. Bibliography

CrowdStrike. (n.d.). Command and control attacks. CrowdStrike. Retrieved from https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/command-and-control-cac-attack/?srsltid=AfmBOoqyWmTqwMlRSKf9d_pnx8HYsVnAuEr5iXxD8Ph880HCQMVcfBHv

Cloudflare. (n.d.). Click fraud botnets. Cloudflare. Retrieved from https://www.cloudflare.com/learning/bots/what-is-ad-fraud/

Cloudflare. (n.d.). DDoS attacks. Cloudflare. Retrieved from https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/

Garcia, S. (n.d.). Malware Capture Facility Project. Stratosphere IPS. Retrieved from https://stratosphereips.org

GeeksforGeeks. (2021, July 21). How to use pandas filter with IQR. GeeksforGeeks. Retrieved from https://www.geeksforgeeks.org/how-to-use-pandas-filter-with-iqr/

Kaggle. (2020, February 13). CTU-13 dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/dhoogla/ctu13

Medium. (2020, November 12). Extracting features from a pcap file and writing to csv/txt using python. Medium. Retrieved from https://medium.com/@ishankarunanayake/extracting-features-from-a-pcap-file-and-writing-to-csv-txt-using-python-c7630ac6322a

Pandas Documentation. (n.d.). DataFrame.describe function. Pandas. Retrieved from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html

ScienceDirect. (2022, August 16). CTU-13 dataset: A comprehensive dataset for network traffic analysis. ScienceDirect. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352340922008344

Stratosphere IPS. (n.d.). CTU-13 dataset. Stratosphere IPS. Retrieved from https://www.stratosphereips.org/datasets-ctu13

Tshark. (n.d.). Tshark manual. Wireshark. Retrieved from https://www.wireshark.org/docs/man-pages/tshark.html

Tshark. (2020, May 12). Tshark extraction. Medium. Retrieved from https://medium.com/@ishankarunanayake/extracting-features-from-a-pcap-file-and-writing-to-csv-txt-using-python-c7630ac6322a