AN INVESTIGATION INTO COLLABORATIVE SCANNERS

AUTHOR: MATYAS KOLLERT – M.KOLLERT@STUDENT.TUDELFT.NL

SUPERVISOR: HARM GRIFFIOEN
RESPONSIBLE PROFESSOR: GEORGIOS SMARAGDAKIS

I. Introduction

Background

- port scanning has been researched extensively since it is a way for attackers to find vulnerabilities [1][2]

Gap in reserach

- lack of recent papers regarding coordination of multiple scanners and their long-term behaviour

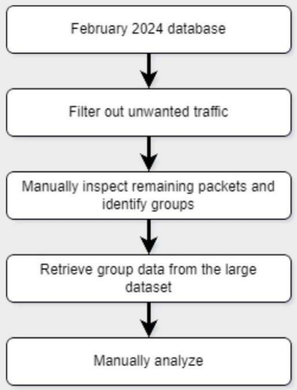
II. Research Question

Aim to manually identify collaborative internet-wide scanners using network telescope data, and describe the patterns of their behavior over time.

- Determine how well can the /24 sub-net, Autonomous System (AS) and temporal patterns in the groups' probing traffic identify collaborative scanners.
- Investigate what trends or changes in the behavior of collaborative scanners can be observed over an extended period of almost a year.

/24 subnet - 192.120.45.11 /0 subnet - whole internet

III. Methodology



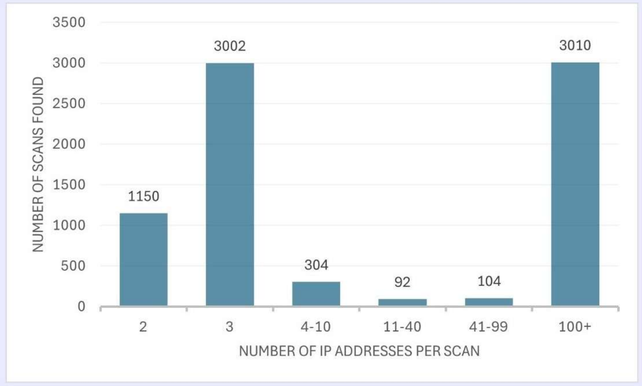
Jupyter Notebooks and Python

- Pandas and Numpy for data manipulation
- Matplotlib for visualization of the findings

IV. Findings

- There are no groups that were identified by the /24 subnet, but not by the AS (all found subnets contained by one AS)

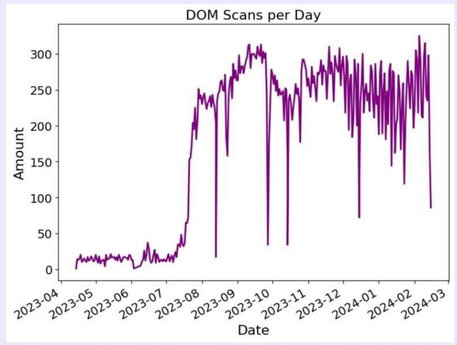
Scan Distribution



- 7662 scans which were then aggregated into 41 groups
  - 10 collaborative scanners described in detail
  - 31 lack enough information
- at least 34 groups coming from a hosting provider
  - 26 from DigitalOcean
  - 3 from Akamai, 2 from CARInet, 2 from Amazon, 1 from Hurricane

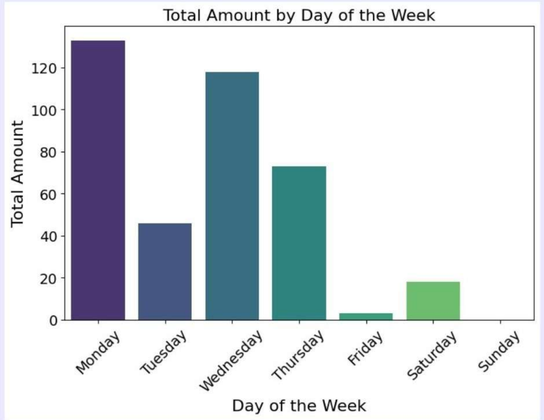
DIGITALOCEAN–MAIN (DOM)

- over 60% of all traffic coming from DigitalOcean
- 3 IP addresses, specific destination partition, rarely seen again



The amount of scans that the DOM collaborative scanned performed per day

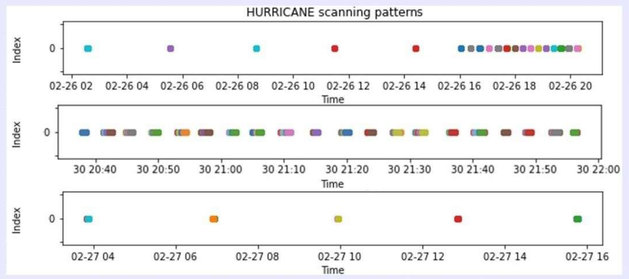
CARINET–01 (CR01)



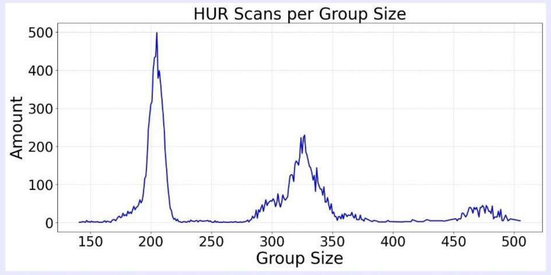
Histogram showing the amount of scans per each day of the week by CR01

HURRICANE (HUR)

- most active group with the most IP addresses active at once



Request timing patterns seen in all scans performed by HUR



Distribution of the amount of IP addresses performing a scan from HUR

V. Conclusion

- Proposed method does detect known scanning groups such as SecurityTrails or the Shadowserver
- Most groups prefer using hosting service providers which do not allow scanning
- 25% of groups were successfully tracked and their behaviour documented
  - Others might be either less relevant or more stealthy
  - Each group had several clear patterns
- Could be easily extended in order to find more sophisticated collaborative scanners

VI. Limitations

- We do not have access to the whole /0 network
- Data is not labeled, therefore we need to define a group ourselves
- Confined to standard ZMap packets
- Only IP addresses that scan in bursts are considered
  - Continuous scanning shows no clear patterns, too much data for manual id
  - Most IP addresses scan in bursts

References

[1] Zakir Durumeric, Michael Bailey, and J. Alex Halderman. An Internet-Wide view of Internet-Wide scanning. In 23rd USENIX Security Symposium (USENIX Security 14), pages 65–78, San Diego, CA, August 2014, USENIX Association.

[2] Harm Griffioen and Christian Doerr. Discovering collaboration: Unveiling slow, distributed scanners based on common header field patterns. In NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, pages 1–9, 2020.