Background image by Microsoft Copilot

Detecting Collaborative ZMap Scans

1 Introduction

- Due to increase in number of connected devices, attackers have a wide range of targets. They often perform internet-wide scans to find vulnerabilities using scanners such as ZMap which are programs used to scan IP addresses for accessible ports.

- Such scans provide malicious parties an overview of potential vulnerabilities and system weaknesses, which they can exploit.

- While it is trivial to detect and block scans from a single device, detecting coordinated scans from multiple devices is significantly harder.

- Detecting distributed scans offers valuable information about the attackers and the specific services they are targeting [1].

2 Research Questions

How can we detect collaborative ZMAP scans in network telescope data using an algorithmic approach?

1. What are the characteristics of collaborative scaning activities using ZMap in network telescope data?

2. How can set cover algorithms be applied to distinguish collaborative scanners from other network traffic?

3. What are the main challenges in detecting collaborative ZMAP scanners and how can they be addressed?

3 Background

- **Internet Scanning:** The process of checking IP addresses on the internet to find open ports, services and potential security gaps. Attackers perform scans to find vulnerabilities in connected devices [1].

- **ZMap Tool:** A high-speed network scanner capable of scanning the entire IPv4 address space under 45 minutes [2].

- **Network Telescopes:** A segment of unused IP addresses that monitors unexpected incoming traffic to detect network security incidents such as DDoS attacks, Internet worm infections and network scanning activities [1, 3].

- **Set Cover Problem:** The challenge of finding the smallest number of subsets from a given collection such that their union is equal to the universal set. It is one of the Karp's 21 NP-complete problems meaning it is a challenging problem known to be difficult to solve efficiently [4, 5].

4 Methodology

Algorithmic Approach

1. Analyzing network telescope data

- The network telescope at TU Delft monitors three IP ranges.
- Data was captured in February 2024 and contains 12.64 billion scans.
- 4.55 billion of these scans were performed using ZMap.
- To reduce noise, only addresses scanned more than 1,000 times were considered.
- Focused on the top four most scanned ports: 8728 (MikroTik RouterOS API), 80 (HTTP), 22 (SSH), and 443 (HTTPS).

2. Analyzing ZMap packets

- Simulated distributed scans using ZMap's --dryrun option.
- The simulation was performed using four sources (shards) on specific IP range and port.
- Examined packet generation and distribution to understand ZMap's scanning behavior.

3. Adapting set cover algorithm

- Developed a new algorithm inspired by the greedy set cover algorithm to detect collaborative scans.
- Steps included:
  1. Define a window size of 1 hour.
  2. Query data within this window.
  3. Create a universal set of destination IP addresses.
  4. Map source IPs to the destination addresses they scanned.
  5. Sort sources by the number of unique destinations they cover.
  6. Iteratively select sources to cover all destinations without overlaps.
  7. Adjust window sizes incrementally to improve detection accuracy.

4. Algorithm validation

- Generated synthetic sources and distributed destination addresses in the universal set with random portions among these sources
- Injected these sources into the dataset to test if the algorithm could detect them.
- Evaluated detection accuracy across different ports (8728, 80, 22, 443).
- Ensured algorithm could detect coordinated scans without false positives.

5. Identifying challenges

- Any challenges encountered during the first four steps of the methodology, along with the solutions implemented to tackle these problems and their effectiveness, are documented.

5 Results

Characteristics of Collaborative ZMap Scanning Activities

- Packets almost evenly distributed across sources.
- Using ZMap sharding to distribute the scan does not result in overlapping scans.
- Noise filtering reduced distinct destination addresses from 174,669 to 62,242 in the network telescope data.

Set Cover Algorithm

- **Validation:** Algorithm tested with 288 injected groups over three days. 260 groups detected, showing >90% accuracy.
- **Port 80:** 8,527 groups identified. Groups from DigitalOcean, Akamai/Linode, SecurityTrails, Gemnet, Rapid7.
- **Port 22:** 212 groups found. Groups from DigitalOcean, Akamai/Linode, Palo Alto Networks, Shadowserver Foundation.
- **Port 443:** 7,331 groups identified. Groups from DigitalOcean, Akamai/Linode, Amazon AWS, Google, Shadowserver Foundation.
- **Port 8728:** Only one group detected. Majority scans done individually, not collaboratively.

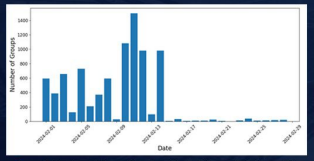

Figure 1: Number of Groups Found per Day in February 2024 (Port 80)

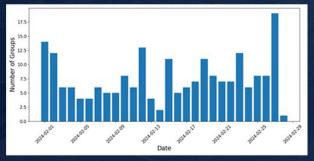

Figure 2: Number of Groups Found per Day in February 2024 (Port 22)

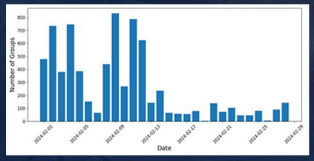

Figure 3: Number of Groups Found per Day in February 2024 (Port 443)

Challenges and Solutions

Challenge 1: Detecting groups that scan multiple times per day. **Solution:** Used shifting windows (1 to 24 hours).
Challenge 2: Long execution time of the algorithm. **Solution:** Parallelized runs per day and port, reducing time to 1.2 hours.
Challenge 3: Identifying actual groups and eliminating mixed sources. **Solution:** No viable solution found.
Challenge 4: Missed groups due to greediness of the algorithm. **Solution:** Variations were just different combinations of the same sources.

6 Conclusions

- Utilizing network telescope data with an adapted greedy set cover algorithm improves the detection of distributed scanning operations using ZMap.

- This approach enhances the ability to identify and mitigate coordinated scanning activities, contributing to improved network security.

- The research demonstrates that algorithmic approaches effectively detect and respond to sophisticated scanning techniques used by attackers.

7 Limitations

- The algorithm relies on specific ZMap scanning characteristics, potentially limiting applicability to other scanning tools.

- Analysis based on a limited dataset from TU Delft's network telescope.

- Algorithm's execution time may increase with larger datasets and complex scanning patterns.

- Unable to detect groups scanning more frequently than once per hour or spanning longer periods.

- Verification of the intent behind identified scanning groups was beyond the study's scope.

8 Future Work

- Optimize algorithm performance for faster scanners, larger datasets and diverse scanning strategies.

- Explore integrating machine learning techniques to enhance detection accuracy.

- Develop methodologies for verifying the objectives of identified scanning groups to improve understanding and response strategies.

References

[1] H. Griffioen, "Scanners: Discovery of distributed slow scanners in telescope data" 2018.

[2] D. Adrian, Z. Durumeric, G. Singh, and J. A. Halderman, "Zippier ZMap: Internet-Wide scanning at 10 gbps," in 8th USENIX Workshop on Offensive Technologies (WOOT 14), (San Diego, CA), USENIX Association, Aug. 2014.

[3] D. Moore, C. Shannon, G. M. Voelker, and S. Savage, "Network telescopes: Technical report," 2004.

[4] J. A. Filar, M. Haythorpe, and R. Taylor, "Linearly-growing reductions of karp's 21 np-complete problems," arXiv preprint arXiv:1902.10349, 2019.

[5] R. M. Karp, Reducibility among combinatorial problems. Springer, 2010.

Author: Fatih Acikkollu - m.f.ackkollu@student.tudelft.nl

TU Delft

Supervisors: Harm Griffioen, Georgios Smaragdakis