State of UDP scanners on the Internet

Sina Rostami
Max Planck Institute for Informatics
Saarbrücken, Germany
srostami@mpi-inf.mpg.de

Taha Albakour
Max Planck Institute for Informatics
Saarbrücken, Germany
talbakou@mpi-inf.mpg.de

Tiago Heinrich
Max Planck Institute for Informatics
Saarbrücken, Germany
theinric@mpi-inf.mpg.de

***Abstract*—Internet scanning is a prevalent event on the Internet. Scanning can have both benevolent and malicious intents. It can potentially be utilized to find vulnerable machines and services to then initiate further intents, such as Denial of Service (DoS) attacks. Therefore, studying the behavior of scanners, the services they are targeting, and the information they are looking for is essential for protecting Internet-facing services and preventing malicious intent against them. In this work, we present our first steps towards our methodology for studying the behavior of Internet scanners. Specifically, we investigate the payloads of the first scan-packets in User Datagram Protocol (UDP). We group payloads based on their similarities to see different characteristics of scanners of the same protocol/service. We find that the variety of scan payloads depends on the protocol for example, 99% of NTP scans have the exact same payload. However, in SIP, this number reduces to 12%. We also highlight the existence of non-standard payloads such as MGLNDD scans in common UDP services.**

***Index Terms*—Internet scan, UDP, clustering, SNMP.**

I. Introduction

Internet scanning is a technique used by researchers and malicious actors to identify and collect information about active hosts and services on the Internet. Scanner attempts to interact with a service by initiating a preparatory contact in a way that the service understands and responds to. The response potential contains useful information that the scanner use to understand the corresponding service or host. Preparatory contact can be as simple as a TCP handshake initiated with a SYN packet —as used widely in common fast scanning tools such as ZMap [1] and Masscan [2]— or more sophisticated approach by speaking the target protocol language and sending specially crafted application layer payload with the goal of eliciting a response. Such approach is often used against UDP based services as they do not require a transport layer handshake. As such, the clients can send their request within the first packet. Scanners can generate payloads that simply check if a service exists, cause denial of service, or even exploit existing vulnerabilities [3].

Several previous studies have looked into the scan traffic characteristics. Hiesgen et al. [4] propose a reactive telescope framework to interact with TCP scanners by completing handshake. They show that a significant share of scans have malicious intents. However, the majority of prior work focus only on studying the behavior of stateless TCP scanners. Ghiette and Doerr [5] develop a method to identify scan campaigns over seven years data of scan traffic by extracting patterns and templates from scan payloads. They analyze DNS and NTP traffic using payload clustering method and show that clustering payload data based on their commonalities can effectively identify scanner behaviors. As an example, they find that DNS scanners have shifted their attention from ANY to TXT query type over time. However, the scope of the work is limited to only DNS and NTP services and their corresponding standard ports. Durumeric et al. [6] highlight DNS, SIP, SNMP, Microsoft SSDP, and CHARGEN services as the most commonly targeted UDP services in their 5.5 million IPv4 addresses telescope. They mention NTP based Denial of Service (DOS) amplification attacks in 2013, in which a Cloudflare customer was targeted with 400 Gbps DDoS attack [7]. They correlate this attack with a large scanning campaign for vulnerable hosts within hours of the Cloudflare incident. Furthermore, they also show that attackers have started using fast scanning tools like ZMap [1]. However, their work focuses mainly on the longitudinal and high level statistics of scanning activities rather than fine-grained in-depth information about UDP scanners behavior.

UDP is widely used in various applications where no reliable connection is required. Applications using UDP are like any other applications prone to bugs and vulnerabilities that could expose the security of the application or the host at risk. For example, existence of a vulnerability in SNMP service [8] on Cisco routers could allow attacker to collect device information, enable backdoor access, and potentially remote code execution on the devices [3].

In this work, we address the gap in prior work and analyze UDP scan payloads towards a telescope machine on the Internet. We present our methodology for studying different scanning behaviors. Based on the similarities of scan payloads, we divide scans into different clusters where each cluster represents a scanning behavior or characteristics. Our methodology reveals the existence of non-standard and unexpected payloads such as IP-in-UDP, Ethernet-in-UDP, and MGLNDD scans targeting various services. Our similarity-based clustering approach also shows that scanners probe for common UDP services on non-standard ports. Further, we find that the scan payloads heterogeneity depends on the protocol being scanned; For example, 99% of NTP scans have the exact same payload while this number reduces to 12% in SIP.

978-3-903176-74-4

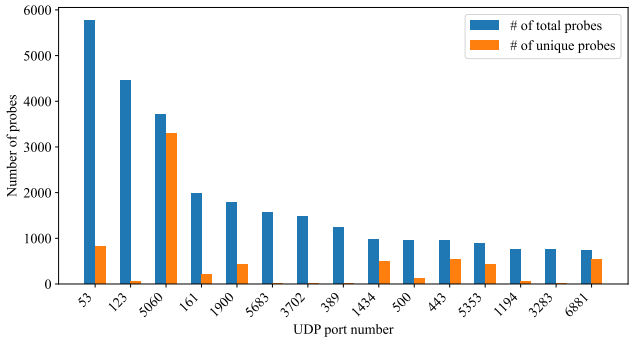

Fig. 1. Number of probes per each of the top 15 UDP port. Orange bars show number of probes with unique payloads.

II. Methodology

Our approach for studying scan payloads and behavior is to investigate their similar characteristics. We extract a 6-tuple for each scan packet from our traffic capture to build our dataset, that is, $\langle time, src\_ip, ttl, src\_port, dst\_port, payload \rangle$. Scan payloads are represented by transformation of the bytes to their respected hexadecimal string notation. Then we form a distance matrix for the dataset records corresponding to their payload distance with each other which is calculated utilizing *Damerau-Levenshtein* edit distance [9]. We then group similar payloads based on this commonality metric.

A. Dataset

We use captured traffic on an Internet facing machine with public IPv4 and IPv6 addresses. For the current study, we use data from the time period between the beginning of December 2024 and end of March 2025.

Only packets with UDP transport protocol are taken into account. Moreover, all local network traffic such as NTP or DNS lookups are filtered out. After data cleaning, the dataset contains over 63K UDP scans packets, which contributes to an average of more than 614 packets per day. Figure 1 shows total and unique number of UDP probes per destination port during the study period. For example, the machine has received more than 4.4K scans to port 123, but only 46 of them have unique payloads while the rest contain duplicated payloads.

B. Clustering

Since there is no prior knowledge about number of clusters, we perform a Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) clustering algorithm and keep track of state of clusters in each step to identify the variety and distribution of the payloads. Dendrogram plot in Figure 2 illustrates an overview of the clusters in varying epsilon values, limited to 10 levels. By inspecting records in several produced clusters, we find that payloads targeting same service are normally more similar to each other and place in the same clusters. Some of these similarities within scan payloads targeting same service are shown as highlighted areas in Figure 2. To validate effectiveness of our clustering approach, we inspect some clusters related to popular services manually and check if the cluster payloads are legitimate and parsable.

III. Preliminary Findings

We analyzed more than 63K UDP scan first packets originating from more than 9.5K unique addresses and containing more than 18K unique payloads.

By analyzing scan clusters, we find that some scanners send payloads belonging to a service to some other ports. Figure 3 shows that while about 25% of popular payloads (among top 30) are seen in scans targeting only 1 port, more than 40% of the payloads are seen to be targeting at least 30 different port numbers.

Following clustering payloads and validating components of clusters containing most popular ports (e.g., DNS, NTP, SIP, SNMP), we then check scans within those clusters targeting other port numbers to see what other ports are being scanned with similar payloads. We find that more than 60% of the formed clusters have scans targeting at least two different port numbers. This number goes higher to at least 5 different port numbers in 20% of the clusters, Highlighting that scanners are scanning non-standard ports with the payloads belonging to some other standard port number.

We take a deeper look at the Simple Network Management Protocol (SNMP) and by analyzing different clusters containing scans to port 161 and then looking into scans targeting other ports within those clusters, we find that SNMP scans are also seen to be sent to 10 other port numbers. These scans could be attempts to conduct UDP port scanning by sending legitimate SNMP payloads to discover running services. Moreover, we describe scan payloads for each cluster containing SNMP scans in Table I.

Looking into clustered payloads, we find that scanners send exact same payload. Therefore, we define a metric to study patterns of common scan payloads. Payloads uniqueness percentage is calculated by dividing number of unique payloads (e.g., eliminating duplicated payloads) by total number of payloads (e.g., duplicated payloads included). In this study, we calculate scan payloads uniqueness for most popular ports with the highest number of payloads. Figure 4 shows an ECDF of this metric for top 30 ports. In NTP scans, only about 1% of scans contain a unique payload, meaning that 99% of them send duplicated payloads. However, in SIP scans, more than 88% of scans contain a unique payload. Number of possible

TABLE I
Content of payloads in each cluster for scans to port 161.

Cluster ID	Cluster Payloads Content
Cluster-1	SNMP version 1/2c *get-request*
Cluster-2	SNMP version 2c *getBulkRequest*
Cluster-3	SNMP version 3 without *msgUserName*
Cluster-4	SNMP version 3 with *msgUserName*
Cluster-5	**MGLNDD** scans
Cluster-6	IP Network Layer packet in UDP payload
Cluster-7	Ethernet Data Link Layer frame in UDP payload

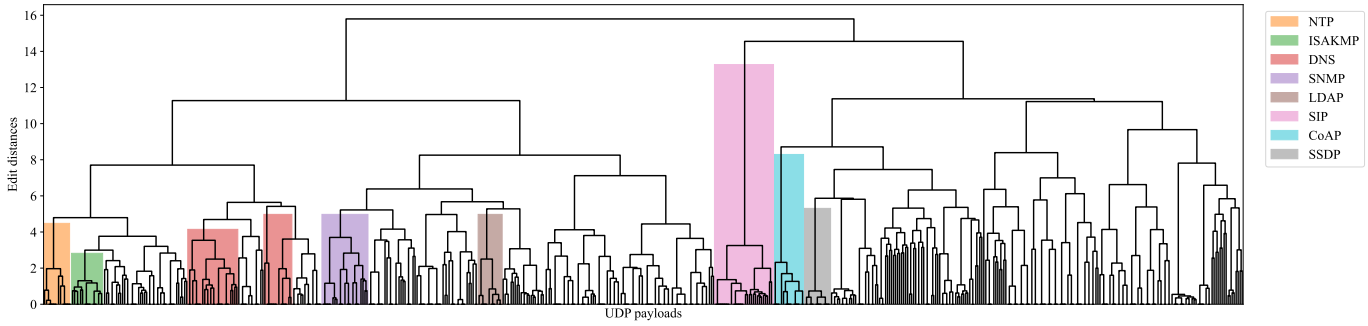

Fig. 2. Dendrogram of a 2K sample payloads based on their similarity, truncated at 10th level. Colors represent samples of detected protocols in the clusters.

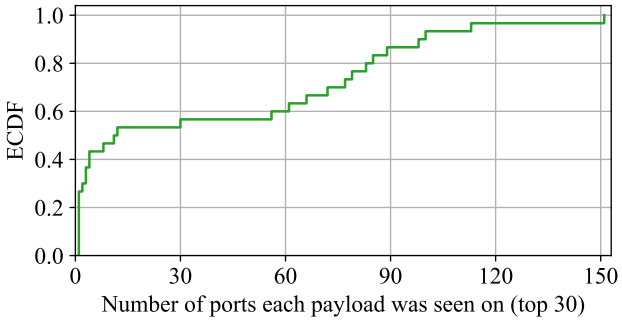

Fig. 3. ECDF plot of Number of ports that each of the top 30 most popular payload patterns were seen on.

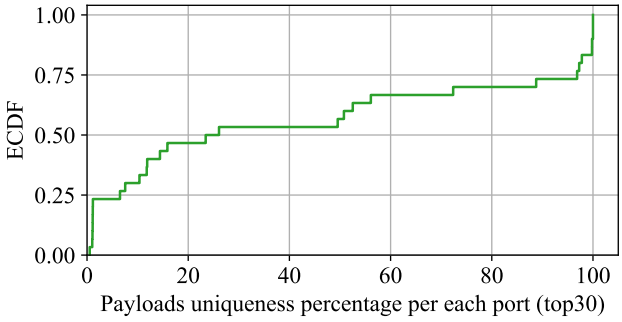

Fig. 4. ECDF of scan payloads uniqueness percentage for top 30 ports.

random fields in the application layer header for each protocol could be a potential factor for this variance.

Other than sending payload of one protocol to a different non-standard port number, we notice non-standard categories of scan payloads as well. We find a type of scan payloads in which the payload is structured based on the host and port number they are targeting. The payload structure is $MGLNDD\_\{IP\}\_\{PORT\}$. From the data in the structure, we believe that it belongs to an asynchronous and stateless scanning framework, in which the scan state information is embedded into the probes to then be restored from their reply. The Internet Storm Center (ISC) suspect that this type of scans originate from RIPE Atlas probes [10]. However, none of the source IPs were found in the database of RIPE Atlas probes [11]. To further explore this behavior, we initiate a RIPE Atlas measurement targeting a controlled host. However, we did not observe any similar payloads.

Another group of unexpected payloads belongs to the encapsulated data. We notice usage of both (IP and Ethernet)-in-UDP encapsulation methods in scanning standard ports, unlike the suggested method in an IETF draft [12].

IV. Conclusion and Future Direction

In this work, we introduce a methodology for studying UDP scanner behaviors and provide insights into ports, services, and resources they probe. We find that the uniqueness of scan payloads varies depending on the protocol being scanned. Our methodology shows scanning of non-standard and unallocated port numbers using payloads from standard protocols. We also highlight existence of unexpected payloads toward standard port numbers, such as MGLNDD scans. To further explore the state and implications of UDP scanners on the Internet, we plan to extend our methodology and attempt to correlate scanning activities with known vulnerabilities.

References

[1] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast internet-wide scanning and its security applications," in *22nd USENIX Security Symposium*, 2013.

[2] R. D. Graham, "Masscan: Mass ip port scanner," 2013, accessed: 25-04-2025. [Online]. Available: https://github.com/robertdavidgraham/masscan

[3] "Jaguar tooth," Apr. 2023, accessed: 25-04-2025. [Online]. Available: https://www.ncsc.gov.uk/static-assets/documents/malware-analysis-reports/jaguar-tooth/NCSC-MAR-Jaguar-Tooth.pdf

[4] R. Hiesgen, M. Nawrocki, A. King, A. Dainotti, T. C. Schmidt, and M. Wählisch, "Spoki: Unveiling a new wave of scanners through a reactive network telescope," in *31st USENIX Security Symposium (USENIX Security 22)*, 2022.

[5] V. Ghiette and C. Doerr, "Clustering payloads: Grouping randomized scan probes into campaign templates," in *2022 IFIP Networking Conference (IFIP Networking)*, 2022.

[6] Z. Durumeric, M. Bailey, and J. A. Halderman, "An internet-wide view of internet-wide scanning," in *Proceedings of the 23rd USENIX Conference on Security Symposium*, 2014.

[7] M. Prince, "Technical Details Behind a 400Gbps NTP Amplification DDoS Attack," 2014, accessed: 25-04-2025.

[8] NIST, "CVE-2017-6742." 2017, accessed: 25-04-2025. [Online]. Available: https://nvd.nist.gov/vuln/detail/cve-2017-6742

[9] G. V. Bard, "Spelling-error tolerant, order-independent pass-phrases via the damerau-levenshtein string-edit distance metric," in *Proceedings of the Fifth Australasian Symposium on ACSW Frontiers - Volume 68*, 2007.

[10] D. Stevens, "Mglndd scans," 2024, accessed: 25-04-2025. [Online]. Available: https://isc.sans.edu/diary/30686

[11] RIPE NCC, "Ripe atlas: A global internet measurement network," 2015, accessed: 25-04-2025. [Online]. Available: https://atlas.ripe.net/

[12] X. Xu, H. Assarpour, S. Ma, D. Bernier, D. Dukes, S. Hegde, Y. Lee, and F. Yongbing, "Encapsulating IP in UDP," Internet Engineering Task Force, Internet-Draft draft-xu-intarea-ip-in-udp-13, 2024, work in Progress. [Online]. Available: https://datatracker.ietf.org/doc/draft-xu-intarea-ip-in-udp/13/