How to Operate a Meta-Telescope in your Spare Time

Daniel Wagner*
DE-CIX
Max Planck Institute for Informatics

Sahil Ashish Ranadive
Georgia Institute of Technology

Harm Griffioen
Delft University of Technology

Michalis Kallitsis
Merit Network, Inc.

Alberto Dainotti
Georgia Institute of Technology

Georgios Smaragdakis
Delft University of Technology

Anja Feldmann
Max Planck Institute for Informatics

ABSTRACT

Unsolicited traffic sent to advertised network space that does not host active services provides insights about misconfigurations as well as potentially malicious activities, including the spread of Botnets, DDoS campaigns, and exploitation of vulnerabilities. Network telescopes have been used for many years to monitor such unsolicited traffic. Unfortunately, they are limited by the available address space for such tasks and, thus, limited to specific geographic and/or network regions.

In this paper, we introduce a novel concept to broadly capture unsolicited Internet traffic, which we call a “meta-telescope”. A meta-telescope is based on the intuition that, with the availability of appropriate vantage points, one can *(i)* infer which address blocks on the Internet are unused and *(ii)* capture traffic towards them—both without having control of such address blocks. From this intuition, we develop and evaluate a methodology for identifying unlikely to be used Internet address space and build a meta-telescope that has very desirable properties, such as broad coverage of dark space both in terms of size and topological placement. Such meta-telescope identifies and captures unsolicited traffic to more than 350k /24 blocks in more than 7k ASes. Through the analysis of background radiation towards these networks, we also highlight that unsolicited traffic differs by destination network/geographic region as well as by network type. Finally, we discuss our experience and challenges when operating a meta-telescope in the wild.

CCS CONCEPTS

• **Security and privacy → Network security**.

KEYWORDS

Network Telescope, Internet Scanning.

ACM Reference Format:
Daniel Wagner, Sahil Ashish Ranadive, Harm Griffioen, Michalis Kallitsis, Alberto Dainotti, Georgios Smaragdakis, and Anja Feldmann. 2023. How to Operate a Meta-Telescope in your Spare Time. In *Proceedings of the 2023 ACM Internet Measurement Conference (IMC ’23), October 24–26, 2023, Montreal, QC, Canada.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3618257.3624831

1 INTRODUCTION

A network telescope, or simply *telescope*, is an infrastructure that passively monitors traffic reaching Internet address space that is not assigned to any hosts but is advertised to the global routing system (i.e., *dark address space*). This traffic is by definition *unsolicited* (also known as Internet background radiation—IBR) and is constituted of an evolving mix of diverse traffic components originating from across the whole Internet [7]. Over the years, researchers have been finding ways to extract insights into various Internet properties and phenomena from IBR, such as, e.g., identifying misconfigurations [7] and large-scale malicious activities [21, 35–37, 46], monitoring Internet connectivity [22], inferring the utilization of the IPv4 space [20], etc.

A telescope operator typically dedicates some of its allocated address space—thus one or a few prefixes—and tends to be limited to one geographic region and a few network locations. This address space has to be owned (or, at least, controlled) and advertised by the telescope operator. However, large and more distributed coverage is highly desirable, since certain IBR traffic components have been shown to be localized (with respect to the destination dark space) [27, 44]. The only known telescope that has been reported to be well distributed (spanning 1,300 networks) is operated by a major content delivery network (CDN) provider and represents a variation of the original concept, since it leverages traffic reaching unused protocol ports on actually used (CDN) servers [44].

In this paper, we introduce a novel concept to broadly capture IBR, which we call a “meta-telescope”. A meta-telescope is based on the intuition that, with the availability of appropriate vantage points, one can *(i)* infer which address blocks on the Internet are unused and *(ii)* capture traffic towards them—both without needing ownership or control of such address blocks. Our key observation is that a significant fraction of the Internet address space is indeed *advertised but unused* [43], i.e., it does not host users, servers, or other network equipment. These properties make such space ideal for monitoring unsolicited traffic destined to it that traverses accessible vantage points. From these intuitions, we develop and evaluate a methodology for identifying unlikely to be used Internet address space and build a “meta-telescope” that has very desirable properties, such as broad coverage of dark space both in terms of size and topological placement.

[*] Also with Saarbrücken Graduate School of Computer Science.

IMC ’23, October 24–26, 2023, Montreal, QC, Canada

ACM ISBN 979-8-4007-0382-9/23/10.
https://doi.org/10.1145/3618257.3624831

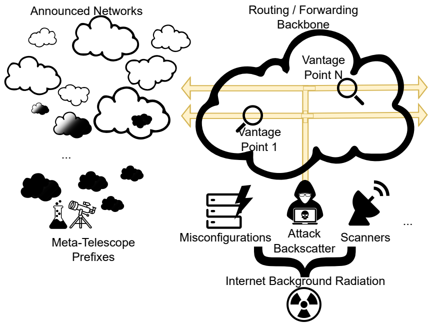


Figure 1: Sketch of the key ideas for identifying potential meta-telescope prefixes.

For an intuitive sketch of these key concepts see Figure 1: Operational telescopes [11, 34] such as the ones shown on the right-hand side of the figure are dedicated network prefixes (black clouds) that are announced for the purpose of attracting IBR (bottom of the figure). The flow of the traffic throughout the Internet is shown as yellow arrows. At network vantage points it is possible to capture such traffic, e.g., as shown at Vantage Point 1. The advantage of monitoring in the middle of the Internet vs. at the edge of the Internet is that one may be able to observe traffic towards topologically and geographically diverse destinations. The disadvantage is that one does not see all traffic towards each destination. On the left-hand side of the figure we show additional network prefixes (visualized as clouds). It is possible to use the vantage points to infer if any of these network prefixes originate any traffic. Using this criterion and auxiliary data sources, we show it is possible to identify potentially-*dark* address blocks within prefixes (black clouds inside the white ones) or even entirely dark prefixes that do not originate any traffic (stand-alone black clouds). These black clouds represent the dark network blocks our meta-telescope monitors to capture IBR.

We propose, implement, and evaluate a pipeline of inference steps to apply to traffic data normally gathered by network monitors in order to identify *candidate* dark address blocks. We refer to them as meta-telescope prefixes. Meta-telescope prefixes can change over time and they are unlikely in the black lists of scanners or attackers. Moreover, the inference of meta-telescope prefixes helps us annotate flows that communicate with inactive address space. This enables us to scale up our ability to infer Internet background radiation as well as narrow down its originating networks at scale. The only requirement for our methodology is access to a network vantage point that regularly captures Internet traffic.
Our contributions can be summarized as follows:

- We develop and evaluate a methodology to identify globally advertised but unused address space to cumulatively contribute to a distributed meta-telescope. We refer to these prefixes as meta-telescope prefixes.
- Meta-telescope prefixes can be identified on demand according to various requirements regarding geographical footprint, network location, and address block size.
- Our analysis, using an Internet exchange point as vantage point, shows that in a single day more than 350k /24 IPv4 prefixes can be identified as meta-telescope prefixes. These are spread across more than 190 countries and 7,000 networks of various types, ranging from "eyeball" to data center to corporate. To the best of our knowledge, the cumulative size of meta-telescope prefixes we detect allows us to build the largest and most distributed telescope up to date.
- We comment on our experience in detecting and operating a meta-telescope in the wild. While spoofing can significantly affect our detection capabilities we show how to overcome this issue.
- We discuss how a meta-telescope can be used to shed light on scanning and other network activity across the Internet and geographical regions, as well as across types of networks.

2 BACKGROUND & RELATED WORK

For more than two decades, network telescope instrumentation—and more broadly, passively capturing and analyzing *unsolicited traffic* (IBR) at various vantage points—have allowed global visibility into a wide range of Internet phenomena: the automated spread of malicious software such as Internet worms or viruses [35–37, 46]; random spoofed source denial-of-service attacks [38]; large-scale botnet activities [21, 42]; macroscopic Internet blackouts due to natural disasters [18], network failures [6] and state censorship [22]; trends in IPv4 address space utilization [19, 20]; bugs and misconfigurations in popular applications [7], etc. Measurement and analysis of such macroscopic phenomena are of key relevance for the security and reliability of the Internet infrastructure.

In the past, researchers have deployed dedicated infrastructure [5, 17, 48] for capturing traffic reaching large unused address blocks. However, due to the increasing scarcity and commercial value of IPv4 address space, the size of even the largest telescopes has been progressively eroding over the years. E.g., this is the case of the two largest telescopes broadly accessible to academic researchers: the UCSD Network Telescope [11]—once almost a /8 block—has been gradually shrinking as subnets are assigned by the owner, and recently saw a whole quarter of its addresses being sold [28]; the Merit Network telescope [34] has also progressively shrunk from originally a /8 to approximately the equivalent of a /13 due to steady sub-allocations.

For these reasons, researchers have also started developing ways to observe unsolicited traffic leveraging existing infrastructure. In 2012, Glatz et al. developed a scheme to dissect *one-way* (thus unsolicited) traffic observed in unsampled NetFlow records from the border routers of a regional academic backbone network. This traffic includes therefore packets destined to active hosts but towards ports where they do not host services or run clients. Unsampled traffic capture at medium-sized (or larger) networks is impractical and the infrastructure used in that study has been discontinued.

In 2019, Richter and Berger presented a "distributed telescope" approach leveraging existing logs of unsolicited packets blocked at the firewalls of ≈ 90,000 servers of a major CDN [45]. These servers are distributed over more than 1,300 networks and are live, offering services to end users. The distributed nature of this setup enabled the authors to uncover phenomena that cannot be captured by individual telescopes placed in only one location of the Internet topology. Specifically, they found evidence of local concentrations of unsolicited traffic, a phenomenon also recently observed by Hiesgen et al. [27] when comparing traffic from telescopes in Europe and the US.

The importance of widely distributed capture approaches and of sensor placement as a key factor in understanding and generalizing measurements from unused address space, were already highlighted in foundational studies from 20 years ago [5, 15, 38]—when large portions of unused IPv4 space were more accessible to researchers. The authors were able to deploy an "Internet Motion Sensor" (a hybrid telescope / honeynet-like infrastructure) consisting of 28 monitored blocks at 18 physical installations. Since then, researchers have rarely had access to largely distributed telescope infrastructure. In contrast, access to largely distributed honeynets is more common, but they represent infrastructure with different goals and characteristics—i.e., targeting specific classes of phenomena (e.g., malware, bruteforcing, exploits, etc.) and with each individual vantage point typically covering only one address or a small block [40].

The original approach we propose aims at *(i)* leveraging existing infrastructure (a meta-telescope operator does not need to own and allocate address space) while *(ii)* enabling coverage that is both broad in size and diverse in terms of topological placement. In addition, *(iii)* our approach offers the opportunity to identify untapped portions of unused IPv4 address space across the whole Internet. This feature also brings *(iv)* another advantage: large individual telescopes tend to become notorious and their address blocks are often blacklisted by scanners and malicious actors [4]; by leveraging uncovered dark address blocks, our meta-telescope also promises to be more resistant to blacklisting. Finally, we expect our approach to bode well with IPv6 traffic monitoring too. However, we leave this exploration to future work, as unsolicited IPv6 traffic has different characteristics in terms of distribution across the IPv6 address space due to its vastness [17].

3 DATA SETS

In this section we describe the vantage points and data sets used in our study. Specifically, Section 3.1 discusses the multiple vantage points we use to infer dark network blocks in the whole public IPv4 Internet and to characterize the IBR they receive. In our study we also use traffic data from three operational telescopes and an operational network that we were granted access to. We analyze specific properties of their traffic to inform several of our parameter choices in our inference methodology such as tuning of thresholds for average packet sizes (see Section 4). We describe these data in Section 3.2. Finally, Section 3.3 lists auxiliary data sets we used such as IP geolocation data or data that allow us to identify with certainty some active (i.e., non dark) subnets.

3.1 Network Vantage Points: IXP Sites

We partner with 14 Internet Exchange Points (IXPs) that have established Internet peering infrastructure in 3 regions of the world, i.e., North America, Central Europe, and South Europe. An IXP is a physical infrastructure offering a logical layer-2 Ethernet switching fabric to its members [13]. Depending on their size, these IXPs are a distributed network over one or multiple peering facilities in a metropolitan area. Network operators can join an IXP to exchange traffic with other members. The IXP's members can choose to place their border routers in the same physical location with IXP switches, or exchange traffic using remote peering [39]. Thus, network operators that are mainly geo-located in other parts of the world can still contribute to the traffic exchanged in an IXP. In some very large IXPs, up to 30% of the members are remote. Typically, an IXP has a mix of contributing network types. Many of the large cloud providers and content delivery networks are members of the IXPs, as well as enterprise networks and regional or national eyeball networks [2, 13]. The size of the IXPs in our study varies in terms of number of members, i.e., peering networks, as well as peak traffic. For an overview of the IXPs in our study see Table 1.

IXP Code	#Members	Peak Traffic (Gbps)	Region	#Sampled Flows (Billions)
CE1	1,000+	12,000+	Central Europe	68.461
CE2	250+	150+	Central Europe	0.904
CE3	200+	150+	Central Europe	0.381
CE4	200+	150+	Central Europe	0.492
NA1	250+	1,000+	North America	8.471
NA2	125+	600+	North America	2.379
NA3	20+	10+	North America	0.031
NA4	20+	50+	North America	0.159
SE1	200+	1,000+	South Europe	2.807
SE2	10+	200+	South Europe	0.978
SE3	40+	50+	South Europe	0.23
SE4	40+	300+	South Europe	1,146
SE5	20+	10+	South Europe	0.179
SE6	30+	15+	South Europe	0.049

Table 1: IXPs: Basic statistics—week of April 24th 2023.

For our study, we get access to network data collected at each of the 14 IXPs. The data is exported through the Internet Protocol Flow Information Export (IPFIX) protocol [14] and contains aggregated packet header information about network flows of the IXPs. It does not contain any packet payloads. The flows are generated on a packet sampling. Collection of the flow data took place in the week from April 24th, 2023 to April 30th, 2023. The total amount of sampled data at all IXPs is 86,667 billion flows estimated to carry about 880 Petabytes of traffic.

3.2 Operational Telescopes

We obtained access to data from three network telescopes that are operated by three different organizations in three different countries. We use these telescopes to inform several of our parameter choices in our methodology (see Section 4), such as tuning of thresholds for average packet sizes. This type of data-driven indicators help us differentiate between "active" versus "dark" /24 networks. We also use these operational data sets to gather insights regarding top-targeted ports. We then juxtapose these observations with the ones obtained through our meta-telescope to shed light into spatial differences in scanning.

We analyze traffic for the week[1] April 24 – April 30, 2023 (see Table 2). We note that the TEU2 telescope only became operational during the course of our study. In addition, the network hosting this telescope is directly peering at ten of the IXPs and has transit connectivity via a tier-1 provider. The largest telescope in our study consists of 1,856 contiguous /24 subnets and each /24 subnet receives an average of 1.91 million packets per day. The share of TCP traffic is 93.8%. Similar results hold for the TEU1 telescope. The TEU2 telescope receives more UDP traffic than the other two and

[1] We verified that we find consistent values for each individual day.

Code	Location	Size (#/24s)	Daily /24 pkt count	Share of TCP traffic	Avg. IP pkt size (TCP)
TUS1	North America	1856	1.91M	93.82%	40.7B
TEU1	Central Europe	768	1.79M	90.38%	40.55B
TEU2	Central Europe	8	2.29M	79.5%	40.78B

Table 2: Operational telescopes: Basic statistics.

its share per /24 is also larger. Two of the telescopes receive traffic on any TCP/UDP port; for the TEU1 telescope ports 23 and 445 are blocked by their ingress router. It is noteworthy that some of the 768 /24 blocks in TEU1 are dynamically allocated to end users on a daily basis so that not all blocks are always actually dark. Table 2 also shows that the total packet count of IBR per /24 does not vary drastically. It typically lies to around 2 million packets per day per /24. (For TEU1 it is less due to some ports being blocked, as mentioned earlier.) We leverage this information in our inference methodology.

To better calibrate our inference approach to distinguish dark vs. active subnets, we also leverage traffic data from the same ISP that hosts the TUS1 telescope. Specifically, we use NetFlow records for the same week of April 24 – April 30, 2023. The ISP receives traffic for 26,079 unique /24 subnets, including those of the TUS1 telescope.

3.3 Auxiliary Datasets

In addition to traffic data from IXP vantage points and operational telescopes/networks, we use a variety of data sets to either validate and supplement our inference pipeline or analyze our results. For a comprehensive statement on ethical considerations, we refer to Section 5.

BGP Routing Data. As a reference and starting point for our inferences, we use the portion of the IPv4 address space that is actually reachable through BGP. To obtain a complete list of prefixes announced in the global routing system, we use routing table (RIB) dumps from a large Route Views collector (route-views4). For each day we consider, we combine the prefixes from all 12 RIB dumps available (Route Views RIBs are dumped every 2 hours). In addition, in the analysis of our results we characterize the portion of inferred dark address space of individual Autonomous Systems (ASes). To this end, we use CAIDA's prefix-to-AS mapping data set (published daily) from April 24, 2023 [9], which is based on Route Views RIB dumps. We also use CAIDA's AS to Organization mapping data set [10] (published every few months) from April 11, 2023. This data set is generated using WHOIS information available from Regional and National Internet Registries to infer a mapping from AS numbers to the organizational entities that operate them.

Data on active IP addresses. To partially validate our results and, afterwards, as a supplemental source of information, we use data from three measurement projects that confirms "liveness" of individual IP addresses: M-Lab's Network Diagnosis Tool (NDT), Censys, and ISI's Internet Address History. The NDT Speed Test is a single-stream performance measurement, initiated by users, of a connection's capacity for "bulk transport". We extract from the NDT "Unified Views" data the list of IPv4 addresses that perform speed tests and mark their /24 IPv4 prefixes as "used" on a daily basis for the week of April 24 – 30, 2023. The Censys Universal Internet [23] data set is generated by scanning the entire IPv4 address space on multiple ports and protocols on a daily basis. We identify roughly 4.8 million active /24 IPv4 prefixes using this data for the week April 24 – April 30, 2023. Finally, the history bits data from Internet Address History data set [47], generated on March 6, 2023, contains IPv4 addresses that respond to ICMP echo requests for scans starting from 2019. We aggregate data for the February, 2023 scan to generate a set of roughly 5.1 million /24 IPv4 prefixes, where we observe each prefix to contain at least 1 address which responded to the ICMP echo requests.

IP Geolocation and AS classification. We geolocate—at a country level—the prefixes that our methodology identifies as advertised through BGP but unused using IP geolocation data from the Maxmind GeoLite2 data set [32] from April 25, 2023. To classify ASes into business categories, we use the IPInfo "IP to Company" commercial database from April 23, 2023.

4 METHODOLOGY

In this section we describe our methodology and inference pipeline for identifying candidate meta-telescope prefixes. Since this study is the first attempt at developing a "meta-telescope" (i.e., a first-of-its-kind analysis) both our filter logic as well as the thresholds are chosen conservatively to ensure that we can identify meta-telescope prefixes with a high confidence while keeping the number of false positives low. Unless stated otherwise, all numbers in this section are for all 14 IXPs and the 24-hour period from April 24th, 2023 00:00 UTC through April 25th, 2023 00:00 UTC.

4.1 Telescope Traffic Analysis

A key intuition of our inference methodology is that traffic towards dark address blocks (IBR) is likely to present different characteristics from traffic towards populated address blocks (i.e., hosting services, clients, etc.). We indeed find that the size of the vast majority of IBR packets tends to be minimal (i.e., typically just made of the IP and TCP header). To derive a packet size threshold to use in our inference pipeline, we perform a detailed analysis of this property by leveraging traffic data from the operational telescopes and the ISP described in Section 3.2.

In the last column of Table 2 we report the average IP packet size we observe for TCP packets captured at each of the three operational telescopes: in all three cases, the average value is smaller than 41 bytes (and greater than the minimum size of 40 bytes, that is, 20 bytes each for IP header and the TCP header respectively). We also verify that this property is consistently present when we separately aggregate packets by /24 block. We find that *at least* 93% of all TCP packets destined to the telescopes have a size of 40 bytes. We note that 40 bytes correspond to a typical TCP-SYN packet, i.e., a packet with no IP or TCP options set nor any payload content. In addition, we see a step at 48 bytes. This is again a typical size of TCP-SYN packet with one option. Large amounts of packets attempting to open a TCP connection are often originated by malware trying to compromise new hosts, malicious actors performing network reconnaissance, or even benign research scanners [3, 24]. Therefore, we speculate that utilizing the TCP packet size as an indicator or "fingerprint" for distinguishing between "dark" and "live" traffic is a useful filtering step in our inference methodology.

To confirm this hypothesis, we look at (both directions of) the traffic traversing a production network containing both dark and active subnets. Specifically, we analyze NetFlow traffic data captured at the border routers of the same ISP that operates the TUS1 telescope (described in Section 3.2). Looking at traffic destined to the ISP, we find 26,079 unique /24 subnets that receive traffic (including the /24 subnets of the TUS1 telescope). On the other hand, only 7,923 /24 subnets out of the 26,079 ones are seen to be originating traffic within that ISP. Hence, for the week of interest, there are 18,151 (i.e., 26,079 - 7,923) *dark subnets*—including the 1,856 /24 subnets dedicated to the TUS1 telescope—that by definition are not used. To identify *active subnets* we focus on the 7,923 networks that we see activity from. To filter out networks that may be considered active because of spoofed traffic, we impose the conservative constraint that a network is considered active only if we have observed at least 10 million packets originating from that subnet during the course of the full week (based on our observation of the distribution of per subnet packet counts). With this constraint at hand, the number of active /24 subnets identified drops to 5,835.

Now that we have obtained "labels" for active and dark subnets, we can select and tune a classification criterion to optimally distinguish between the two classes. For this work, we evaluated two features (see Table 3): (i) median packet size and (ii) average packet size destined to a /24 subnet. The classification rule simply checks whether the median (average) packet size destined at a /24 subnet is less or equal than a threshold of N bytes. If so, that /24 subnet is considered to be "dark". Otherwise, it is classified as "active". We experimented with various threshold choices for N, and the sensitivity analysis of Table 3 showcases our results. A good classification outcome is reached when using the *average packet size* criterion with a threshold of 44 bytes. This setting achieves very high accuracy and, importantly, a very low false positive rate. This value provides the second-best F1-score[2]. Nevertheless, between the two best thresholds (44 versus 46 bytes average packet size), we opt to use the 44 bytes threshold due to its lower false positive rate.

4.2 Inference Pipeline

Our inference pipeline consists of seven steps (Figure 2) to exclude address blocks that are unfit (e.g., reserved space) or whose traffic does not exhibit typical IBR characteristics. At a high level, we seek the following characteristics for the meta-telescope prefixes: (a) they are routed and advertised but not reserved for special purposes [16]; (b) they do not originate any traffic; (c) the size of their incoming TCP packets is small, i.e., with an average smaller than 44 bytes (recall 3.2); and (d) they do not receive too much traffic.

We start with the roughly 6 million IPv4 destination /24 subnets that are included in the IXP traffic data set and apply the following filtering steps:

1. TCP Traffic. As noted earlier, TCP SYN packets are very common in IBR and UDP is very noisy. As such, we remove any subnet that does not receive any TCP traffic. About 300k /24 subnets are filtered out in this step.

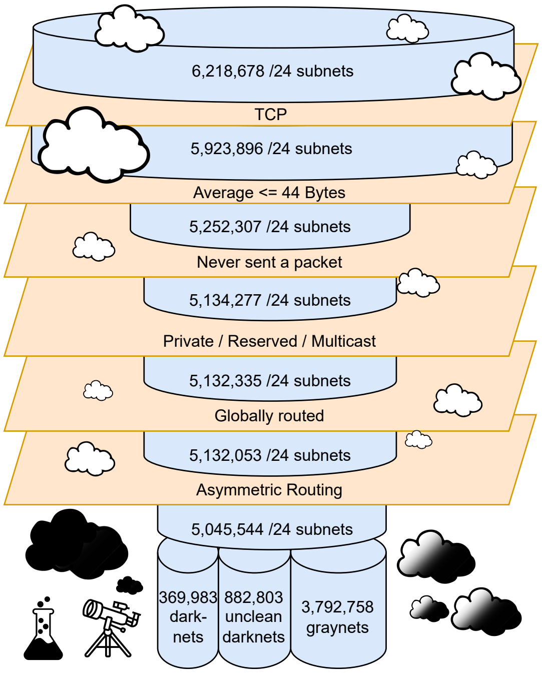


Figure 2: Illustration of inference pipeline: # of /24 blocks—all IXPs and 24th April 2023.

2. Average Packet Size. Moreover, as TCP SYN packets dominate background radiation TCP traffic, we remove any subnet that receives TCP traffic with an average packet size above 44 bytes. This filter removes about 800k /24 subnets from the data set.

3. Source Address Unseen. Next, we require that potential meta-telescope prefixes do not originate any traffic. Hence, we remove any subnets from which we observe traffic. Some 100k subnets are removed in this step. While this is a straightforward filter, it is susceptible to source address spoofing, thus also removing subnets that do not actually generate traffic (but whose addresses were used in spoofed packets). Yet, the likelihood that the sampled IXP traffic includes scanning packet towards an address and in the same period spoofed packets "from" that address is rather low. However, as we increase the duration of the data set, this issue has to be addressed, since the likelihood of including spoofed packets increases.

4. Private / Multicast / Reserved. Telescopes must be reachable in the public Internet. Hence, we remove all IP blocks from the IXP data set that are inside private IPv4 blocks, multicast IPv4 blocks, or reserved IPv4 blocks. About 2 thousand /24 subnets are removed by this filter.

5. Globally Routed. Our next filter ensures that candidate dark prefixes should not only be located inside a routed address block, but also need to be publicly announced. Here, we rely on daily snapshots from Route Views [1] and remove any /24 subnet that is not inside an announced prefix as seen by Route Views. About 300 subnets are hereby removed.

6. Asymmetric Routes. Another challenge is due to asymmetric routing in the Internet. For example, Content Delivery Networks (CDNs) often receive lots of TCP ACK packets from an IP, but send their data to them via another path not visible at any of the IXPs.

[2] The F1-score measures the overall classification accuracy, and is defined as $F_1 := 2\text{tp}/(2\text{tp} + \text{fp} + \text{fn})$, where tp denotes the true positives, fp the false positives and fn the false negative results.

Median Packet Size	Classified Dark, but are Active	Classified Active, but are Dark	Classified Dark, and are Dark	Classified Active, and are Active	F1-score
Threshold (bytes)	False Positive Rate	False Negative Rate	True Positive Rate	True Negative Rate	
40	6.96%	0.39%	99.61%	93.04%	98.70%
42	8.00%	0.39%	99.61%	92.00%	98.54%
44	22.59%	0.09%	99.91%	77.41%	96.45%
46	22.64%	0.09%	99.91%	77.36%	96.44%

Average Packet Size	Classified Dark, but are Active	Classified Active, but are Dark	Classified Dark, and are Dark	Classified Active, and are Active	F1-score
Threshold (bytes)	False Positive Rate	False Negative Rate	True Positive Rate	True Negative Rate	
40	0.00%	99.10%	0.90%	100.00%	1.83%
42	0.53%	56.83%	43.17%	99.47%	60.25%
44	**0.87%**	**0.41%**	**99.59%**	**99.13%**	**99.65%**
46	1.08%	0.27%	99.73%	98.92%	99.69%

Table 3: Tuning the packet-size based fingerprint that allows distinguishing between "active" versus "dark" /24 subnets using ISP data that hosts both TUS1 and active blocks.

These packets will typically be 40 bytes long similarly to the TCP SYN packets targeted by our average-packet-size filter. We leverage the fact that IBR is limited compared to production traffic [41] and apply a conservative volume-based filter of 1.7M packets on average per day per /24. This step filters out 90k /24 blocks.

7. Classification. Finally, we classify all /24 blocks into three classes: (a) dark (i.e., meta-telescope prefix), (b) unclean darknets, and (c) graynets. For a block of IP addresses to be a meta-telescope prefix, all IPv4 addresses have to survive the above filter steps. Unclean blocks are those that have at least one IP surviving the filter steps and at least one IP that did not survive the filters but did not originate traffic. If one IP inside a block of IPs originates traffic in which another IP survived our filters, we consider this a graynet.

After the last filtering step, we are left with 5M /24 blocks. Our classification pipeline labels about 370k of them as dark (potential meta-telescope prefixes), 880k as unclean, and 3,8M as gray.

Code	Size (/24s)	#Inferred meta-telescope prefixes			
		1 day		7 days	
		CE1	All	CE1	All
TUS1	1856	0	437	0	1424
TEU1	768	38	33	262	247
TEU2	8	0	0	7	7

Table 4: Meta-telescope coverage of IPv4 address space of the operational telescopes for 1 and 7 days.

4.3 Evaluation

We evaluate the results of our inference in three ways: *(i)* We verify its ability to identify the address space used by known telescopes; *(ii)* We compare port count statistics from the traffic we observe towards our inferred dark prefixes against traffic observed at operational telescopes; *(iii)* We use three data sets of host activity/responsiveness to identify (a lower bound for) false positives in our inferences. In addition, we use results from this last comparison to further refine and finalize our list of meta-telescope prefixes.

To assess the effectiveness of our inferences, we check if we can infer as dark the address space of the three operational telescopes we have been granted access to (see Section 3.2). Table 4 summarizes our findings. Using 7 days of data from CE1 we can infer as dark 31.15% and 87.5% of the address space of TEU1 and TEU2, respectively. The remaining address space is inferred as either unclean or gray due to spoofing or we did not see any traffic to it. We cannot find the address space of the TUS1 telescope, as it is not visible at this IXP. However, using data from all IXPs we can infer as dark 23.5% of TUS1's address space in a single day and 77% of it in a week. Notably, TEU1 contained 503 active /24 blocks on the day that we report on and, thus, the 38 inferred blocks correspond to 14.3% of the unused space in that telescope.

Overall, we find that our conservative approach of using a threshold of 1.7M packets per /24 per day eliminates quite some /24s of the telescopes. Indeed, it might not necessarily be the ideal choice, as TEU2 receives more than 2.2M packets on average. Furthermore, since the telescope's address space is directly announced at 10 of the vantage points, its traffic is well observed. That is why we are unable to infer a single /24 inside TEU2 even with data for several days. Still, even with this conservative threshold, we can infer a large amount of dark space while not having to deal with many false positives (as we show at the end of this section).

To further underline our ability to detect potentially-dark address space, in Figure 3 we plot a Hilbert curve of an IPv4 address block which contains the address blocks of a network telescope. Every pixel corresponds to a /24 block of the IPv4 space. The colorized pixels correspond to inferred dark blocks. Uncolored pixels correspond to blocks without data or that are inferred as unclean or gray. The boundaries of the operational telescope are marked in gray and we clearly see that almost all blue pixels correctly fall within this area. That there are a few, i.e., 5, outside is also not surprising, since not all of the address space outside this block must necessarily be in use.

We then compare port statistics of packets towards the inferred dark prefixes against those from operational telescopes. Table 5 shows statistics we extract for the top-10 TCP ports by analyzing raw PCAP data collected from the three telescopes. We observe a high degree of similarity between the three sites but also some notable differences. Ports 22, 80, and 443 are in the top list of all three. However, port 6379, a top-5 port in TUS1 and TEU2, does not appear in TEU1. This highlights the importance of studying scanning traffic destined to multiple, distributed vantage points, a significant advantage of our meta-telescope approach. When comparing these numbers against those from traffic we observe at the IXP vantage points towards the blocks we infer as dark, we find a perfect overlap for the top ports, namely 22, 23, 80, 443, and 8080.

The above results confirm our ability to identify dark blocks (true positives) and suggest that overall the traffic towards these prefixes is consistent with high-level properties of IBR. To complete our evaluation, we lastly assess the presence of false positives (i.e.,

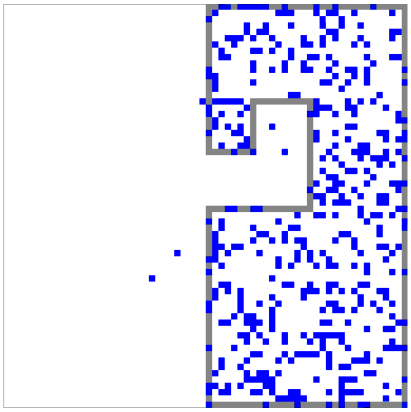


Figure 3: Hilbert curve of IPv4 address space colored by meta-telescope annotated with a gray box for the address space of an operational telescope.

Port Rank	Telescopes		
	TUS1	TEU1	TEU2
#1	23	22	23
#2	6379	80	22
#3	22	443	80
#4	80	8080	6379
#5	443	3389	445
#6	8080	5555	25565
#7	25565	60023	443
#8	5555	81	8080
#9	3389	8443	8090
#10	60023	2375	3389

Table 5: Operational telescopes: Top 10 ports in descending order by popularity (April 24-30, 2023).

blocks inferred as dark whereas they are active) to the extent that publicly available data sets allow for. The indications of activity we use are: *(a)* hosts replying on any transport port upon contacting as reported by Censys [23], *(b)* end-users performing speed tests of their Internet connectivity as reported by NDT [33], and *(c)* hosts replying to ICMP echo requests as reported by ISI [47]. We aggregate this data at /24-block granularity and we compare it with our initial list of 369,983 blocks we infer as dark. We find that 51,337 out of these 369,983 /24 blocks have been active, i.e., 13.9% of our inferred-dark /24 blocks are (at least for some IP addresses) active. This (positive) result shows there is still significant room for improvement to our pipeline in terms of false positives[3]. However, we can apply such active networks ground-truth data to further filter our inferences and obtain a more accurate final set of meta-telescope prefixes. In the remainder of the paper our analysis is based on this final set of prefixes. Table 6 summarizes our results in terms of inferred dark /24 blocks after applying this final correction (by individual vantage points and overall).

4.4 Limitations

The IXP vantage points used in our study impose some limitations to our results.

Sampling. The IXP data set is generated based on packet samples. To gain statistical significance a lot of packets may be required, which is one reason for some of the variations in our results, e.g., when looking at data from different days. Indeed, high-volume DDoS attacks are much easier to capture than low volume background radiation. However, the large amount of addresses being scanned and the possibility to extend our inference to arbitrarily long time frames, help overcome this limitation. We further discuss the impact of sampling in Section 7.3

Routing. Another limitation is that we can only see traffic routed via the vantage point. Note that IXP customers may use alternative routes for part of their traffic, which, for example, restricts the visibility of some of the smaller IXPs.

Asymmetric Routing. Given the prevalence of asymmetric routing, one cannot presume that the observation of traffic in one direction implies that the traffic in the reverse direction is routed over the same path. Hence an IP address block that appears dark, may not be so, as the traffic might be routed via a different path. We tackle this challenge in two ways. First, we only consider networks that receive less than 1.7M packets per day. Second, we eliminate all IPs that appear to be active in any of our auxiliary data sets.

Locality. Given that IXPs pursue the motto: “keep local data local” [13], the data set may be prone to geographical bias. However, (especially) the larger IXPs offer many services that are attractive also to peers from other regions, including ease of connectivity and remote peering [12]. Moreover, hypergiants [8], including cloud providers, often peer at IXPs.

Spoofing. Unfortunately, spoofing—sending packets with an incorrect source IP address—is quite common in the Internet [30]. Spoofing impacts our methodology, since these packets may use source IP addresses of potential meta-telescope prefixes and, thus, will disqualify the whole block. As we expand our data set, e.g., by expanding the observation period or by adding more vantage points, we likely look at more spoofed traffic. As such, the amount of inferred meta-telescope space decreases. To counteract this behavior, we will allow for a small number of potentially spoofed packets, see Section 7.

5 ETHICAL CONSIDERATIONS

Traffic Captures and Data Products. Our study is based on traffic statistics and data that the IXPs may gather for operational purposes and are in compliance with legal requirements in the respective countries of operation. All traffic traces are aggregated at the flow level and do not contain any payload. Additionally, the data is processed and analyzed in-situ at the IXP premises and all analyses can potentially happen in an online manner using only aggregate information at the /24 subnet. We utilize the flow data available at the IXP to extract two data products: (a) the list of network /24 prefixes that, using the methodology proposed in this paper, are inferred to be “unused”—we refer to the set of these prefixes as the meta-telescope; (b) traffic flow traces (which are a subset of the overall IXP flow data sets) destined to the meta-telescope's prefixes that could be used to shed light into Internet-wide scanning activities, malware campaigns, etc.

Leveraging Inferred Unused Prefixes for a Meta-telescope. The key idea of our study is to identify prefixes that can serve as telescopes for a vantage point. One may argue that these prefixes are not controlled by the vantage point operator (i.e., the IXP in

[3] Especially considering that these three data sets can only provide a lower bound for active networks, i.e., they do not necessarily identify all /24 blocks on the Internet that have at least one active IPv4 address.

our case). However, the traffic passes the vantage point and uses the resources of the vantage point. Therefore, this traffic does not differ from any of the other traffic at the vantage point and the vantage point operator has an operational interest in monitoring such traffic. E.g., it is not atypical for operators to use appliances that process the sampled traffic data for upstream analysis, traffic insights, DDoS mitigation and other threat analysis. These appliances monitor traffic that transits the IXPs, which do not own either end of the communication, and offer insights that could be used for traffic engineering purposes, capacity planning, peering arrangements, routing policies, etc. These insights help the operators operate their networks in a safe manner and offer quality service to their customers. This paper proposes a new approach that fits into the same scope of data analysis and allows the vantage point operators to leverage the traffic traces they already process to obtain a novel product, namely the meta-telescope and its data products. These new data products can then be utilized to identify "unwanted" traffic, i.e., background radiation traffic that is inherently malicious/suspicious, since destined to subnets that are seemingly unused. Once the meta-telescope prefixes are identified, only a small number of IP /24 subnets needs to be further monitored (about 5%).

Note that the distilled data products are not meant to be broadly or publicly shared. Although the vantage point operators may elect to share aggregate meta-data (e.g., targeted ports per region or per country) with the community, the main utility of the extracted data sets lies within the operator itself. For instance, the operator may utilize the data sets to enhance the cybersecurity posture of their customers by informing them of suspicious IPs originating from the customer prefixes and destined to the inferred meta-telescope prefixes. Similarly, the meta-telescope operator may share the threat intelligence extracted from the meta-telescope prefixes with appropriate authorities, e.g., Computer Emergency Response Team (CERT) organizations, to inform them about the onset of new malicious activities or nefarious scanning campaigns. To minimize any potential reputation harm that may arise from Type I errors (i.e., false positives) when detecting unused subnets, we followed a very conservative approach in the calibration of the parameters of our methodology (as we had discussed earlier). To further minimize these risks, one could complement the meta-telescope insights with the observations from operational/real network telescopes (such as the three we employ in our study) and other existing threat intelligence data sets.

Operational Telescopes. Our study is based on data that the operational telescopes regularly capture for operational purposes and are again in compliance with the legal requirements in the respective countries. All data is processed and analyzed in-situ at the premise of the telescope operator. The telescopes receive only unidirectional traffic destined to the unused address space, and we do not probe or interact with any of the source IPs sending traffic to the telescope.

Routing and Active Measurement Datasets. Our study is based on data made accessible by the respective data provider either via research license or via public domain access. None of the used data was created specifically for the purpose of this study.

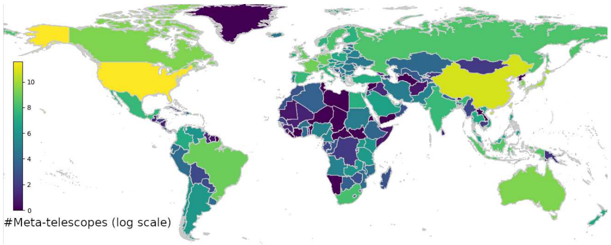


Figure 4: World map colored according to the number of /24 blocks in meta-telescope prefixes (logarithmic scale).

IXP	#Inferred meta-telescope prefixes	#ASes	#Countries
CE1	397,000	8,529	201
CE2	21,340	1,597	124
CE3	61,607	3,982	173
CE4	2,178	455	84
NA1	395,585	8,960	198
NA2	12,489	919	102
NA3	262	128	17
NA4	1,054	299	74
SE1	34,222	2,269	152
SE2	56,638	2,078	132
SE3	3,782	729	97
SE4	43,573	2,431	152
SE5	1,949	667	104
SE6	270	104	33
All	318,646	7,195	194

Table 6: Overview of the meta-telescope prefixes we identify (individual vantage point and overall).

6 META-TELESCOPE PROPERTIES

In this section, we analyze the properties of the meta-telescope we built. For this analysis, we use the ipinfo and pfx2as datasets (see Section 3) to determine the geographic location of the meta-telescope prefixes and to which ASes they belong. This analysis helps us understand the basic properties of the meta-telescope as well as obtain insights about the geographic distribution of the inferred prefixes and their network types.

6.1 Basic Properties

Using our methodology we are able to identify a very large number of /24 blocks as meta-telescope prefixes, namely up to 318,646 /24s in a single day using all vantage points. These are originated by more than 7,000 ASes, which are in turn located in almost 200 different countries (see Table 6).

We find variance in the results depending on the vantage point. Even though some of the vantage points have a very limited visibility on Internet traffic, they are still useful in helping us to identify more than 250 meta-telescope prefixes in various regions of the world while larger vantage points are able to infer thousands of meta-telescope prefixes around the globe. Note that, when combining multiple vantage points, we obtain a smaller number of prefixes compared to the largest individual contributors (CE1 and NA1). This is because more information about individual /24s is provided to our filtering criteria, which are also designed to be conservative and prioritize low false positives.

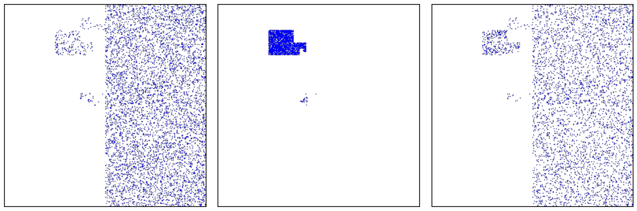

(a) CE1 (b) NA1 (c) All sites

Figure 5: Hilbert maps of a /8 with colored meta-telescope prefixes.

Using the inferred meta-telescope prefixes from all IXPs, in Figure 4 we show a visualization of the distribution of the meta-telescope prefixes across the world in a logarithmic scale. For comparison, we show the maps for IXPs CE1, NA1, and all IXPs in Figure 13-15 in the Appendix. We see that the inferred meta-telescope prefixes are located in IP spaces belonging to almost every country. This includes even small countries that we usually do not have network telescope insights for, given that none of the operational telescopes that we are aware have presence in such countries.

Most meta-telescope prefixes are located in the USA according to our geographic mapping. We find that NA1 which has the "best" visibility of the traffic in that region is able to infer the largest number of meta-telescope prefixes in the USA. By adding information from other vantage points, the number of inferred meta-telescope prefixes in the USA decreases slightly. This is likely due to the impact of spoofed traffic. That the USA is dominating is likely due to the fact that a large fraction of address space was allocated to US-organizations in the early days of the Internet. Most of these large, i.e., mainly /8 blocks, seem to remain primarily unused. To our surprise, the country ranked second by the number of inferred meta-telescope prefixes is China. Visibility into prefixes within China is one example that highlights the benefits of the proposed methodology, since this can allow researchers to study scanning and other macroscopic activities destined to usually unobserved address space. Regions that are still not well covered include central Africa, some middle eastern regions, and North Korea. To overcome this aspect, one might need vantage points closer to these regions.

6.2 Meta-telescope prefixes: Examples

Next, we look at some example meta-telescope prefixes. The first is about a /9 block; the second is about a large network telescope subnet we are aware of.

In Figure 5, we plot the Hilbert map of a /8 where an inferred /9 meta-telescope prefix is inferred using data from three different vantage points, namely (a) CE1, (b) NA1, and (c) all vantage points combined. Here, colored pixels refer to /24 address blocks of inferred meta-telescope prefixes, whereas white pixels correspond to /24 blocks that are either gray or unclean or for which we have no data. We find that CE1 has great visibility of the /9 block in the right half of the Hilbert map. However, the visibility of the left half is not as good. This is partially due to it containing some unannounced space and partially due to lack of traffic.

The NA1 vantage point does not infer a single /24 block of the right /9 block as meta-telescope prefixes. However, it identifies the blocks inside the left /9 as meta-telescope prefixes. These coincide precisely with the ones from CE1. When combining the data from all vantage points, the /9 is again visible, however with a slightly lower density. The latter is likely due to the added noise of spoofing which causes some /24 address blocks to be classified as gray. Note, we do not know whether the /9 on the right side or the /14 on the left side are operational telescopes or if they just happen to be unused IPv4 space.

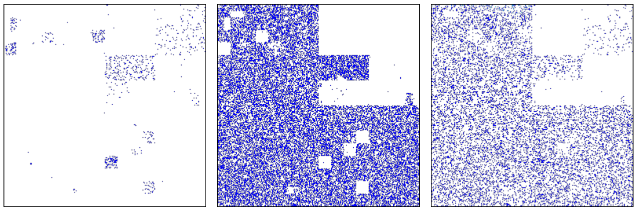

(a) CE1 (b) NA1 (c) All sites

Figure 6: Hilbert maps of a /8 which contains a known telescope with colored meta-telescope prefixes.

World Region	Total	ISP	Enterprise	Education	Data Center
All	318,559	158,262	56,598	79,206	24,493
North America	119,919	30,756	28,407	44,729	16,027
South America	10,680	9,492	849	99	240
Europe	58,990	34,284	11,105	10,323	3,278
Asia	106,411	68,180	12,255	21,958	4,018
Africa	9,411	7,458	1,401	318	234
Oceania	12,373	7,726	2,437	1,741	469
International	729	344	125	38	222

Table 7: Number of meta-telescope /24 prefixes using the union data set, per type and continent.

Next, we take a closer look at another /8. We selected it since we know that it contains an operational telescope. Figure 6 shows the corresponding Hilbert maps in which we observe the address space of the telescope in the upper and lower left, and the lower right quarters. The upper right quarter does not belong to the telescope. Unlike before, CE1 is unable to infer many of the actual telescope prefixes as meta-telescope prefixes, hence the number of colored pixels in Figure 6(a) is small. In Figure 6(b), we show the meta-telescope prefixes inferred by vantage point NA1. Here, we find that many meta-telescope prefixes are inferred correctly and we can clearly see the boundaries of the telescope's address space.

Still, some sizable blocks inside the telescope space remain undetected (white) using data from NA1. However, for some of these blocks, the CE1 vantage point has visibility. When integrating data from all vantage points, our inferred meta-telescope prefixes match the known operational telescope prefixes, see Figure 6(c). One reason for the need of multiple vantage points has to do with the different route announcements, i.e., using more specifics, and preferences of specific neighbors. This results in different route propagation which can lead to some blind spots at certain vantage points, e.g., NA1 or CE1. This example highlights that large as well as small vantage points can add substantial value to the overall visibility. Furthermore, by combining data from multiple vantage points one can increase visibility, but the trade-off is that the spoofing concern becomes more challenging.

6.3 Meta-telescope prefixes: Network Types

Next, we focus on studying the network type of the ASes that host the inferred meta-telescope prefixes. We rely on the classification

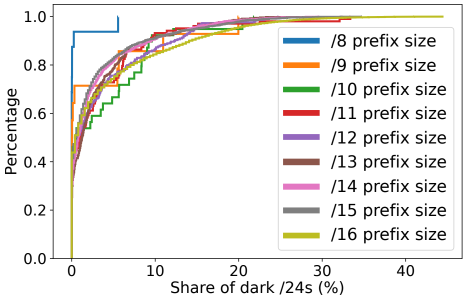


Figure 7: Prefix index: ECDF for different prefix sizes.

offered by the "ipinfo" data set, and consider the following network types: ISP, Enterprise, Education, and Data Center. Table 7 shows the resulting numbers of meta-telescope prefixes per network type and geographic region for all vantage points. This highlights that we are able to identify meta-telescope prefixes in all network types in every region of the world, ranging from a few dozens to tens of thousands. Thus, our methodology gives us access to meta-telescope prefixes inside ISP, Enterprise, and Data Center networks which up to now has been a rather difficult task for researchers. This result underlines the novel contribution of our methodology.

The results confirm that North America hosts the largest share of inferred meta-telescope prefixes. In addition, most meta-telescope prefixes are located inside ISP networks rather than educational networks as is common for most operational telescope networks. The second largest number of meta-telescope prefixes is located in address ranges of educational institutions. The third most ones are inside Enterprise networks, while the smallest number belongs to Data Center networks. Africa is not covered as well. Somewhat surprisingly, this is also true for South America. However, the likely explanation is that we do not have an IXP vantage point within South America. The row labeled as "International" refers to prefixes that cannot be mapped to only one region (a small number of prefixes).

6.4 Meta-telescope: Prefix Coverage

Next, we examine what fraction of an address space is inferred as meta-telescope space regardless of the AS that hosts it. For this task, we focus on large prefixes that are advertised, and can be observed via the Route Views data set. Specifically, we look at advertised prefixes ranging from /8 to /16 and calculate for each the *prefix index*, i.e., the portion of /24 blocks that are identified as meta-telescope prefixes within their covering prefix. In Figure 7 we show an ECDF of the prefix index for each prefix size under consideration.

We find that a surprisingly large share of meta-telescope /24 prefixes is found in such large announced prefixes. For example, more than 6.6% of all /8 BGP announcements have more than 5% meta-telescope address space. This number even rises to roughly 20% for 12% of all /9 announcements. For smaller announced prefixes, i.e., /10 to /15, we find that the share of meta-telescope address space decreases slightly to roughly 10% for most of them. However, we find that for a few /16 announcements, the share of meta-telescope address space is larger than 40%.

We also check if the fraction of meta-telescope space changes with network type, see Figure 16 in the Appendix. We find that there are only small differences, with one exception: data center networks tend to have a smaller fraction of meta-telescope space. This is likely due to the fact that data centers have emerged in times where IPv4 address space was already relatively scarce. Looking by continent (see Figure 17 in the Appendix), EU followed by AF have the least share, which is again consistent with IPv4 address scarcity.

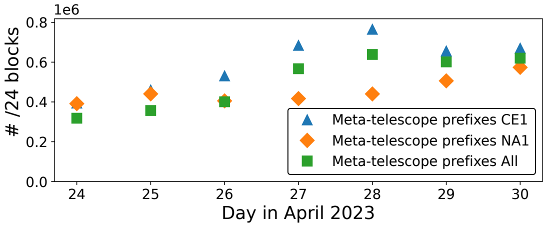


Figure 8: Number of daily meta-telescope prefixes for CE1, NA1, and all.

7 META-TELESCOPE CHALLENGES

Next, we comment on the challenges we encountered while aiming to infer meta-telescope prefixes. The first challenge relates to (packet-based) flow sampling, which results in a high variability in the results depending on the day and the vantage point. The second challenge relates to spoofing, which can significantly impact the results. The third challenge relates to sampling. Lastly, we comment on challenges based on our experience in operating a meta-telescope.

7.1 Meta-telescope Prefixes Variability

To highlight the diurnal variability, Figure 8 depicts the number meta-telescope prefixes for each of the seven days considered in our study. Using data from the vantage point CE1 from April 24th, 2023, we are able to infer 397,000 /24 blocks. In contrast, using data from the same vantage point four days later, we infer roughly twice as many meta-telescope prefixes. The same kind of variability can be observed for the other vantage points during the week. Another trend across all vantage points is that we are able to infer more meta-telescope prefixes during weekends. One possible explanation is that enterprise or educational networks do not have any major activity outside of the working hours. Hence, the effect of sampling becomes less apparent due to the lower traffic volume processed at each vantage point.

Overall, one has to consider multiple trade-offs when inferring meta-telescope prefixes. Should one aim at obtaining stable prefixes, it is better to check if a prefix is among the meta-telescope prefixes identified in multiple days. Moreover, we recommend employing the approach of inferring meta-telescope prefixes on a daily basis also to account for the dynamic nature of Internet routing and address space utilization. Namely, routing changes and/or changes in the use of the address space can alter the observed behavior of network prefixes.

7.2 Effect of Spoofing

Spoofing refers to packets with fake source IP addresses. Thus, if we encounter a packet with a spoofed source IP address from a possible meta-telescope prefix range, we may erroneously classify

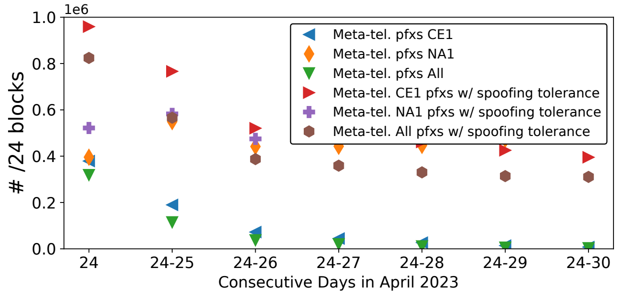


Figure 9: The effect of spoofing to the number of meta-telescope prefixes for CE1, NA1, and All.

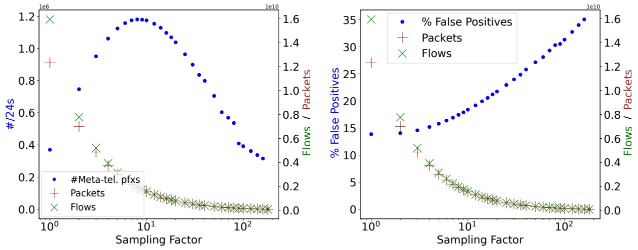
it as an invalid meta-telescope prefix range. Note that spoofing is quite common in the Internet and we have validated that at all of our vantage points we regularly observe spoofed traffic from the three operational telescopes we had access to. We are certain that traffic originating from any operational telescope is spoofed traffic, since these "dark" IP spaces do not initiate, interact with, nor respond to any network traffic.

Spoofed packets negatively effect our inference capability for identifying meta-telescope prefixes. Per our methodology in Section 4, (a) we filter out any destination address that is seen sending traffic, and (b) we classify any /24 block as a graynet if it contains at least one "sending" address. In fact, due to the latter step, even a few spoofed packets from a single address could remove a full /24 block from our candidate meta-telescope prefixes.

Given that actors that resort to spoofing activities are typically selecting IP sources across routed and unrouted address space we can leverage this characteristic to better cope with spoofing [29]. The key intuition here is that one could observe the spoofing activities that occur within *unrouted* IP space, and use this information to obtain a "baseline" for spoofing behavior (similarly to [19]). We examined traffic from known unrouted IP space to identify how many spoofed packets one should expect and adjust the filters in our methodology. As a result, we manage to obtain a tolerance for packets to be seen sourced "from" a /24 block without mistakenly tagging that block as a graynet. We calculate this tolerance for each vantage point and each time frame. Hence, we can adapt to events that may cause an increase in the number of spoofed packets. Concretely, we calculate the 99.99th percentile of packets seen per /24 block inside 2 unrouted /8 blocks per day. We allow this many packets to be sourced by any meta-telescope prefix on that day. For most sites and a single day this threshold is zero packets; for some other sites, it is one or two packets. For seven days the threshold can rise up to four packets per day.

To highlight the impact of spoofing, Figure 9 shows how the number of meta-telescope prefixes that we infer decreases when we add more days of data. The number of blocks for all sites decreases from 350k to 4k. If we add the spoofing tolerance the numbers change from above 800k to 400k. Similar observations hold for CE1.

The vantage point NA1 does not seem to suffer that much from spoofing as the number of inferred meta-telescope prefixes does not decrease substantially. In fact, for some days the inferred meta-telescope prefixes increase. Nevertheless, the spoofing tolerance (which is very small for NA1) is still meaningful and allows us to reach higher numbers of inferred meta-telescope prefixes.



(a) Number of /24 meta-telescope prefixes

(b) Percent of false positives

Figure 10: Effect of performing the meta-telescope prefix inference on various sub-sampled data sets.

7.3 Effect of Sampling

Since our analysis is applied to a data set consisting of sampled data, we want to better understand the impact of sampling to our inference method. To this end, we create additional data sets with different sub-sampling rates from the original data set of all 14 IXPs from April 24th, 2023. As it is not possible to get data from the IXPs using lower sampling rates, i.e., inspecting more packets to create the flow data, we use higher sampling rates, i.e., investigating a lower number of packets, to get a view on the effect of sampling. For a sub-sampling factor of 2, we only consider every second packet in the IXP data. For a factor of 3, only every third, and so on. Figure 10 shows the results when applying our inference method to various sub-sampled portions, starting with the original data set, e.g., a sampling factor of 1. We plot the number of flows and number of packets in the respective sub-sampled data sets. In Figure 10(a), we plot the absolute number of inferred /24 meta-telescope prefixes on the sub-sampled data sets. We see that using higher sampling rates, the inference method first identifies more /24 meta-telescope prefixes, which is likely linked to spoofing becoming less represented in the data set and, hence, eliminating fewer /24 blocks that are actually meta-telescope prefixes. However, when considering only every 100th packet of the original data set, the inference method becomes blind to many parts of the Internet and starts to infer fewer meta-telescope prefixes. Finally, when reaching a sub-sampling rate of 180, our inference method can no longer identify a single meta-telescope prefix. In Figure 10(b), we plot the share of false-positives of the analyses performed on the respective sub-sample data sets. We find that the rate of false-positives is monotonously increasing when using higher sub-sampling rates. Overall, there seems to be a sweet-spot of sampling rates to use, if the data set is prone to spoofing. However, we conclude that using lower sampling rates will provide the most reliable data set of meta-telescope prefixes.

7.4 Summary of Challenges

Overall, under a reasonable sampling rate, false positives (i.e., identifying an active prefix as meta-telescope prefix) are by design not a concern, as we elaborated in Section 4. However, when the sampling rate is very high, false positives are a concern, since there are not enough samples to conclude with high confidence which part of the address space is not active. By increasing the period of our study, see Section 7.1, and due to spoofing, see Section 7.2,

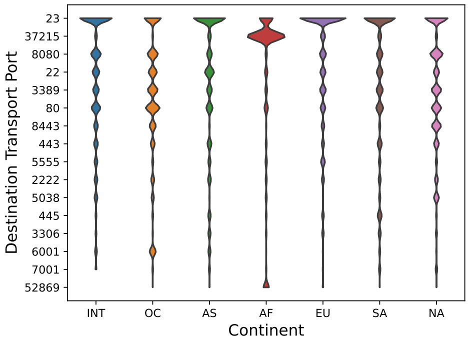


Figure 11: Bean plot of activity for the top 16 destination ports in meta-telescope traffic per world region.

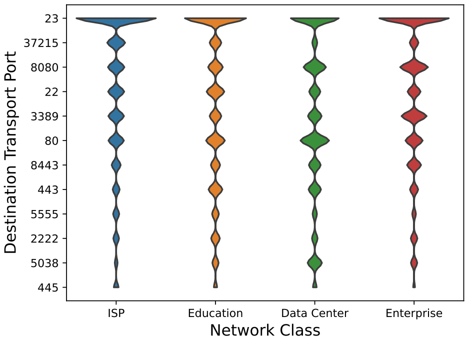


Figure 12: Bean plot of activity for the top 12 destination ports in meta-telescope traffic per network type.

the false negative rate may increase, i.e., non-active prefixes may not be identified as meta-telescope prefixes. The size of vantage points also plays an important role. As we elaborated in Section 6.1, smaller vantage points typically have a higher false negative rate, i.e., less prefixes are identified as meta-telescope prefixes, compared to larger vantage points. The vantage point's physical proximity to the allocated unused space region also plays a role. Typically, the false negative rate is lower for allocated unused space in the same region (continent) of a vantage point, as shown in Section 6.

8 META-TELESCOPE INSIGHTS

In this section, we discuss some insights distilled by the inferred meta-telescope to underscore the new abilities unleashed by this new measurement approach. Specifically, we examine scanning activities targeting different geographic regions and different network types.

8.1 Targeted Ports by Geographic Region

Using the inferred meta-telescope prefixes, we study traffic destined to TCP ports to shed light into services / applications contacted by nefarious actors. This analysis offers insights to cybersecurity analysts about ongoing security incidents, such as, e.g., scans for exploitable ports (e.g., ssh or telnet ports targeted by Mirai botnet variants [4]), reconnaissance activities for vulnerabilities on ports such as 3389 (Microsoft remote desktop services), randomly spoofed DDoS attacks (observed in our inferred meta-telescope as "backscatter" traffic [38]).

We start by inspecting the distribution of the most popular destination ports. We first compile the list of top-targeted ports for all meta-telescope prefixes of each region. We then join these lists to get an aggregate list of top-16 ports. This list includes ports often probed by scanners, e.g., ports 23 (telnet) and port 2222 (used by derivatives of the Mirai botnet [4]), and popular services, e.g., port 80 / 443 (HTTP / HTTPS).

In Figure 11 we use bean plots to visualize the distribution of traffic per top-16 ports as observed by the meta-telescope prefixes. A bean plot allows for visual comparison of univariate data between groups. Here, the groups are the world regions and the values are the port popularity. The ports are ordered by the total popularity in descending order. Overall, we find that port 23 dominates in all regions except OC and AF. Port 37215 is among the top-10 due to its popularity in AF. This port is used for attacks on Huawai HC532 routers. This port is not aimed so aggressively in other regions. Another port that is mainly popular in AF is 52869, which seems to be associated with "Satori", i.e., a Mirai variant that scans both ports 37215 and 52869 to spread. So traffic to port 37215 seems to serve two purposes but most likely it is associated with the Satori botnet. We also see popular Web ports, namely ports 8080, 80, 443, 8443. However, somewhat surprisingly, 8080 is observed to be the most popular one. However, given that network administrators are more prone to securing services listening to TCP/80, adversaries may have adjusted their strategies to look for alternative HTTP ports. We also note that 8080 is another port targeted by the Mirai botnet. Figure 11 shows the port activities relative to the total activity within the specific region (see Appendix C for traffic share relative to the overall traffic).

8.2 Targeted Ports by Network Type

Next, we check if there are interesting observations when we categorize traffic by the network type that the meta-telescope prefix is located in. We again compile top-lists per network type and then obtain their union. This yields a total of 12 top destination ports, see Figure 12. Compared with the lists of the previous section, ports 52869, 6001, 7001, and 3306 are now excluded. No new ports are identified. The ports amiss are ports that are only popular in some of the regions; e.g., port 3306 is mainly popular in AF and NA, port 52869 in AF, port 6001 in OC, and 7001 in NA. Port 23 is again the most popular one. Port 37215 is not that prevalent anymore in a single group, but has contributions in all sub-categories. Still, ISPs contribute the most and these are mainly located in AF. Overall, we see that the popularity of the top 6–8 ports after removing port 23 is roughly similar. Also, note that the port rankings can significantly vary across categories; e.g., port 80 is more popular within data centers and educational networks, and less preferred when it comes to ISP-based meta-telescopes. A likely explanation is that scanners are trying to find unprotected Web servers within data centers. In a similar manner, activity against port 5038—a MLDB database port—is higher within data centers compared to ISP-based or enterprise networks. In Appendix D we provide details on port activity for different network types in NA and EU respectively.

9 DISCUSSION

Dealing with Spoofing. Our experience in detecting meta-telescope prefixes shows that spoofing has a significant effect. Indeed, the number of detected meta-telescope prefixes reduces over time. Even with the spoofing tolerance technique, some missed meta-telescope prefixes can be detected. An alternative approach is to exclude network flows by networks that do not implement spoofing filters, namely BCP 38 [25]. Projects such as Spoofer [30] maintain a list of networks that have not adopted BCP 38. Another approach is to exclude flows originating from IPs that do not belong to a peering network's network cone [31] from our analysis network. We are aware that the network cone is not always accurate. However, this approach will reduce the spoofed traffic significantly. As part of the future work, we would like to study how successful each of these two techniques are in detecting a higher number of meta-telescope prefixes.

The Vantage Point Effect. Our methodology can be applied to network flows collected at any vantage point. For this study, we presented an in-detail analysis of the meta-telescope prefixes that can be inferred when using Internet exchange points of different size in terms of number of members and traffic and from different regions. However, shortcomings such as routing visibility, asymmetric routing, sampled traffic, spoofing, and members' demographics may limit the number of meta-telescope prefixes we discover. Network flows captured at large Internet service providers do not suffer from asymmetric routing. In this case, the routing visibility is less of a concern, and many of them already have BCP 38 implemented. Some ISPs also collect network flow information at a high sampling rate. For all the above reasons, there is the potential to detect even a higher number of meta-telescope prefixes when analyzing ISP data. We plan to apply our method with ISPs in our future work.

Meta-telescope Information as a Service. Internet exchange points and service providers that implement our methodology can detect information about meta-telescope prefixes, and they can also infer which of their peers or customers send traffic to them. They can offer this information as an opt-in service to their customers to make them aware that traffic that originates from their network has a meta-telescope prefix as destination, helping their customers to make traffic engineering and filtering decisions. Our results show that the set of meta-telescope prefixes is quite stable for a couple of days. However, the set of meta-telescope prefixes will vary when the observation window increases in duration and traffic conditions change rapidly, e.g., unused space is allocated to hosts. We argue that additional vantage points and regular measurements to detect meta-telescope prefixes are needed before meta-telescope information as a service will be able to handle spontaneous prefix allocation and traffic changes.

Federated Meta-telescopes. The detection of meta-telescope prefixes can inform and improve the operation of more customers than those of a single IXP or ISP. The detection can be shared among trusted parties to detect meta-telescope prefixes with higher accuracy collectively. It is also possible to develop a standard to enable operators to opt-in to the measurements, e.g., a BGP community or embedment into RPKI that marks announced but unused space. Encoding known only to involved parties will help in keeping this tagging hidden so the prefixes will not be excluded from scanners and attackers. Getting this information into operation has the potential to impact network operators' operations significantly. The research community can also benefit and contribute to this effort.

IPv6 Meta-telescopes. In this paper, we focus on the detection of IPv4 meta-telescope prefixes. IPv6 address space is much larger and less active. In addition, IPv6 address assignment varies per network and vendor. As IPv6 traffic increases, it is important also to detect meta-telescope IPv6 prefixes. Given the vastness of the IPv6 space, our filtering pipeline would likely need adjustments. The lack of complete and reliable hit lists [26] and archives of active measurements for IPv6 further complicate the detection of meta-telescope IPv6 prefixes. We plan to address these issues in future work.

10 CONCLUSION

For decades network telescopes have been used to collect and analyze unsolicited traffic to detect misconfigurations or malicious activities including the spread of Botnets, DDoS campaigns, and exploitation of vulnerabilities. A limitation when deploying a network telescope is that its visibility is limited to the scanning and attack activity received in the announced prefixes, which are typically limited in one regional or geographic location. In this paper, we develop and evaluate a methodology to detect advertised but unused space worldwide and potentially originated by any organization. This approach can provide insights similar to those of collectively operating network telescopes in all portions of this space, which we refer to as meta-telescope prefixes.

By utilizing traffic flows collected in the core of the Internet, i.e., at Internet exchange points, we show that it is possible to detect more than 350k /24 IPv4 address blocks, in 7k ASes and 190 countries, which can be used in a meta-telescope. The size of this meta-telescope is by far larger than any other operational telescope, and it is also highly distributed and observing (unsolicited) traffic towards networks of different types. These features allow us to answer measurement questions for unsolicited traffic arriving in networks at different regions and types, without the need to own, advertise, and dedicate address space for a telescope, and the operational overhead of running it. We also comment on our experience on detecting meta-telescope prefixes. While spoofing can significantly reduce the detection of meta-telescope prefixes, we propose ways to overcome this issue. Our future work includes plans to apply our methodology to other types of vantage points, e.g., large and small ISPs, and compare the obtained meta-telescope datasets with those obtained through IXP vantage points.

ACKNOWLEDGMENTS

We would like to thank our shepherd, Nguyen Phong Hoang, and the anonymous reviewers for their valuable comments. The work of Daniel Wagner was partially funded by the Federal Ministry of Education and Research of Germany (BMBF, grants 5G-INSEL 16KIS0691 and AIDOS 16KIS0975K). The research of Harm Griffioen and Georgios Smaragdakis was supported in part by the European Research Council (ERC) Starting Grant ResolutioNet (ERC-StG-679158). The work of Michalis Kallitsis was partially supported by the NSF CNS award 2120400.

REFERENCES

[1] 2023. Routeviews Project – University of Oregon. http://www.routeviews.org/. (2023).

[2] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. 2012. Anatomy of a Large European IXP. In *Proc. ACM SIGCOMM*.

[3] Aniket Anand, Michalis Kallitsis, Jackson Sippe, and Alberto Dainotti. 2023. Aggressive Internet-Wide Scanners: Network Impact and Longitudinal Characterization. In *ACM CoNEXT*.

[4] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. 2017. Understanding the Mirai Botnet. In *USENIX Security Symposium*.

[5] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson. 2005. The Internet Motion Sensor - A Distributed Blackhole Monitoring System. In *NDSS*.

[6] Karyn Benson, Alberto Dainotti, kc Claffy, and Emile Aben. 2012. Gaining Insight into AS-level Outages through Analysis of Internet Background Radiation. In *ACM CoNEXT Student Workshop*.

[7] Karyn. Benson, Alberto. Dainotti, kc. Claffy, Alex C. Snoeren, and Michalis Kallitsis. 2015. Leveraging Internet Background Radiation for Opportunistic Network Analysis. In *ACM IMC*.

[8] Timm Böttger, Felix Cuadrado, and Steve Uhlig. 2018. Looking for Hypergiants in PeeringDB. *ACM SIGCOMM Computer Communication Review* 48, 3 (2018).

[9] CAIDA. 2023. Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6 . https://www.caida.org/catalog/datasets/routeviews-prefix2as/. (2023).

[10] CAIDA. 2023. The CAIDA UCSD AS to Organization Mapping Dataset, 2023-04-11. https://www.caida.org/data/as_organizations/. (2023).

[11] CAIDA. 2023. The UCSD Network Telescope. https://www.caida.org/projects/network_telescope/. (2023).

[12] Ignocio Castro, Juan Camilo Cardona, Sergey Gorinsky, and Pierre Francois. 2014. Remote Peering: More Peering without Internet Flattening. In *Proc. ACM CoNEXT*.

[13] Nikolaos Chatzis, Georgios Smaragdakis, Anja Feldmann, and Walter Willinger. 2013. There is More to IXPs than Meets the Eye. *ACM SIGCOMM Computer Communication Review* 45, 5 (2013).

[14] Benoit Claise, Brian. Trammell, and Paul Aitken. 2013. RFC 7011: Specification of the IPFIX Protocol for the Exchange of Flow Information. (2013).

[15] Evan Cooke, Michael Bailey, Z. Morley Mao, David Watson, Farnam Jahanian, and Danny McPherson. 2004. Toward Understanding Distributed Blackhole Placement. In *Proceedings of the 2004 ACM Workshop on Rapid Malcode (WORM '04)*. Association for Computing Machinery, New York, NY, USA, 54–64. https://doi.org/10.1145/1029618.1029627

[16] Michelle Cotton, Leo Vegoda, Ronald P. Bonica, and Brian Haberman. 2013. Special-Purpose IP Address Registries. IETF RFC 6890. (2013).

[17] Jakub Czyz, Kyle Lady, Sam G. Miller, Michael Bailey, Michael Kallitsis, and Manish Karir. 2013. Understanding IPv6 Internet Background Radiation. In *ACM IMC*.

[18] Alberto Dainotti, Roman Amman, Emile Aben, and Kimberly C. Claffy. 2012. Extracting Benefit from Harm: Using Malware Pollution to Analyze the Impact of Political and Geophysical Events on the Internet. *ACM SIGCOMM Computer Communication Review* 42, 1 (2012), 31–39.

[19] Alberto Dainotti, Karyn Benson, Alistair King, kc Claffy, M Kallitsis, E Glatz, and X Dimitropoulos. 2014. Estimating Internet Address Space Usage Through Passive Measurements. *ACM SIGCOMM Computer Communication Review* 44, 1 (2014).

[20] Alberto Dainotti, Karyn Benson, Alistair King, Bradley Huffaker, Eduard Glatz, Xenofontas Dimitropoulos, Philipp Richter, Alessandro Finamore, and Alex C. Snoeren. 2016. Lost in Space: Improving Inference of IPv4 Address Space Utilization. *IEEE Journal on Selected Areas in Communications (JSAC)* 34, 6 (2016), 1862–1876.

[21] Alberto Dainotti, Alistair King, kc Claffy, Ferdinando Papale, and Antonio Pescapè. 2012. Analysis of a "/0" stealth scan from a botnet. In *ACM IMC*.

[22] Alberto Dainotti, Claudio Squarcella, Emilie Aben, kc Claffy, Marco Chiesa, Michele Russo, and Antonio Pescape. 2011. Analysis of Country-wide Internet Outages Caused by Censorship. In *ACM IMC*.

[23] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning.

[24] Zakir Durumeric, Michael Bailey, and Alex J. Halderman. 2014. An Internet-Wide View of Internet-Wide Scanning. In *USENIX Security Symposium*.

[25] Paul Ferguson and Daniel Senie. 2000. Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing. IETF RFC 2827. (2000).

[26] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczynski, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *ACM IMC*.

[27] Raphael Hiesgen, Marcin Nawrocki, Alistair King, Alberto Dainotti, Thomas C Schmidt, and Matthias Wählisch. 2022. Spoki: Unveiling a New Wave of Scanners through a Reactive Network Telescope. In *USENIX Security Symposium*.

[28] Kantor, Brian and Karn, Phil and Claffy, kc. and Gilmore, John and Magnuski, Hank and Garbee, Bdale and Hansen, Skip and Horne, Bill and Ricketts, John and Traschewski, Jann and Vixie, Paul. 2019. AMPRNet. https://web.archive.org/web/20190719144558/https://www.ampr.org/amprnet/. (2019).

[29] Franziska Lichtblau, Florian Streibelt, Thorben Krüger, Philipp Richter, and Anja Feldmann. 2017. Detection, Classification, and Analysis of Inter-Domain Traffic with Spoofed Source IP Addresses. In *Proceedings of ACM IMC 2017*.

[30] Matthew Luckie, Robert Beverly, Ryan Koga, Ken Keys, Joshua A. Kroll, and k claffy. 2019. Network Hygiene, Incentives, and Regulation: Deployment of Source Address Validation in the Internet.

[31] Matthew Luckie, Bradley Huffaker, kc Claffy, Amogh Dhamdhere, and Vasileios Giotsas. 2013. AS Relationships, Customer Cones, and Validation. In *ACM IMC*.

[32] Maxmind GeoLite2. 2023. GeoIP2 and GeoLite City and Country Databases. https://www.maxmind.com. (2023).

[33] Measurement Lab. (2023-04-24 – 2023-04-30). The M-Lab NDT Data Set. https://measurementlab.net/tests/ndt. ((2023-04-24 – 2023-04-30)).

[34] Merit Network, Inc. 2023. ORION: Observatory for Cyber-Risk Insights and Outages of Networks. https://www.merit.edu/initiatives/orion-network-telescope/. (2023).

[35] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. 2003. Inside the Slammer Worm. *IEEE Security and Privacy* 1, 4 (2003), 33–39.

[36] David Moore and Colleen Shannon. 2005. The Spread of the Witty Worm. *IEEE Security and Privacy* 2, 4 (2005), 46–50.

[37] David Moore, Colleen Shannon, and kc Claffy. 2002. Code-Red: A Case Study on the Spread and Victims of an Internet Worm. In *ACM Internet Measurement Workshop*.

[38] David Moore, Geoffrey Voelker, and Stefan Savage. 2001. Inferring Internet Denial-of-Service Activity. In *USENIX Security Symposium*.

[39] George Nomikos, Vasileios Kotronis, Pavlos Sermpezis, Petros Gigis, Lefteris Manassakis, Christoph Dietzel, Stavros Konstantaras, Xenofontas Dimitropoulos, and Vasileios Giotsas. 2018. O Peer, Where Art Thou? Uncovering Remote Peering Interconnections at IXPs. In *ACM IMC*.

[40] Honeynet Project. 2002. Know Your Enemy: Honeynets. (Nov 2002). http://projects.honeynet.org/papers/honeynet/.

[41] Lin Quan, John Heidemann, and Yuri Pradkin. 2013. Trinocular: Understanding Internet Reliability Through Adaptive Probing. In *Proc. ACM SIGCOMM*.

[42] Elias Raftopoulos, Eduard Glatz, Xenofontas Dimitropoulos, and Alberto Dainotti. 2015. How Dangerous Is Internet Scanning? A Measurement Study of the Aftermath of an Internet-Wide Scan. In *Traffic Monitoring and Analysis Workshop*.

[43] Philipp Richter, Mark Allman, Randy Bush, and Vern Paxson. 2015. A Primer on IPv4 Scarcity. *ACM SIGCOMM Computer Communication Review* 45, 2 (Apr 2015).

[44] Philipp Richter and Arthur Berger. 2019. Scanning the scanners: Sensing the internet from a massively distributed network telescope. In *Proceedings of the Internet Measurement Conference*. 144–157.

[45] Philipp Richter and Arthur Berger. 2019. Scanning the Scanners: Sensing the Internet from a Massively Sistributed Network Telescope. In *ACM IMC*.

[46] Stuart Staniford, David Moore, Vern Paxson, and Nicholas Weaver. 2004. The Top Speed of Flash Worms. In *ACM Workshop on Rapid Malcode (WORM)*.

[47] USC/LANDER project. 2023. Internet Addresses IPv4 Response History Dataset. https://ant.isi.edu/datasets/index.html. (2023).

[48] Vinod Yegneswaran, Paul Barford, and Dave Plonka. 2004. On the Design and Use of Internet Sinks for Network Abuse Monitoring. In *Recent Advances in Intrusion Detection*, Erland Jonsson, Alfonso Valdes, and Magnus Almgren (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 146–165.

APPENDIX

A WORLD MAPS

For comparison, we also plot the world maps indicating the number of inferred meta-telescope prefixes in logarithmic color scheme using data from vantage point CE1 in Figure 13, NA1 in Figure 14 and all vantage points combined in Figure 15. We note the different color scaling for NA1 as the number of inferred meta-telescope prefixes is higher for this vantage point. We find that all vantage points infer meta-telescope prefixes in all regions of the world. Once again, most inferred meta-telescope prefixes are located in the USA. Vantage point CE1 has the best visibility of meta-telescope prefixes that belong to China.

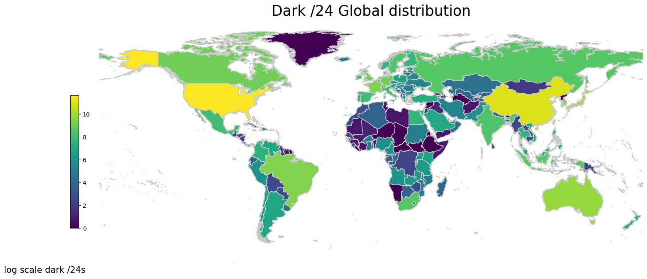


Figure 13: World map colored according to the number of /24s meta-telescope blocks as seen by `CE1`(logarithmic scale).

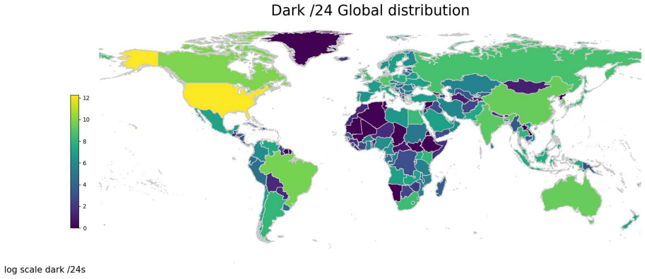


Figure 14: World map colored according to the number of /24s meta-telescope blocks as seen by `NA1`(logarithmic scale).

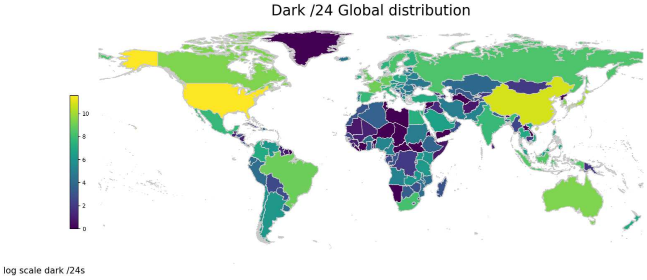


Figure 15: World map colored according to the number of /24s meta-telescope blocks as seen by all sites combined (logarithmic scale).

B META-TELESCOPE PREFIX DISTRIBUTION

We also check if the fraction of meta-telescope space changes with network type. We find that there are only small differences with one exception. Figure 16 shows that Data Center networks tend to have a smaller fraction of meta-telescope space. Looking by continent, see Figure 17, EU followed by AF have the least share which is again consistent with IPv4 address scarcity.

C TARGETED PORTS BY REGION

Following up on the insights from Section 8.1, we plot the port activities in all regions relative to the overall traffic share in all meta-telescope prefixes in Figure 18. It highlights that SA, OC, and INT receive a very small share of the overall traffic. Figure 18 also puts into global perspective the activity against port 37215 which is observed to dominate the AF region.

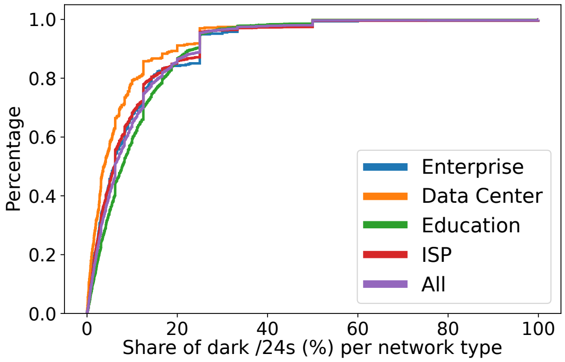


Figure 16: Network type index: ECDF for different network types.

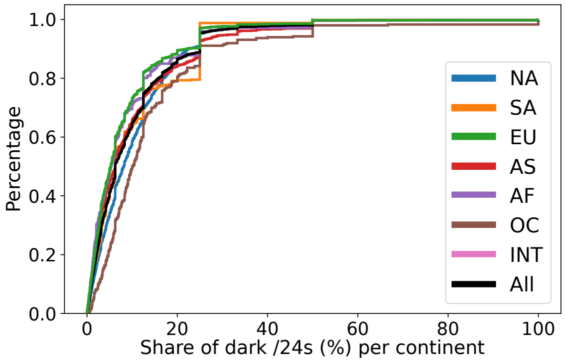


Figure 17: Continent index: ECDF for different world regions.

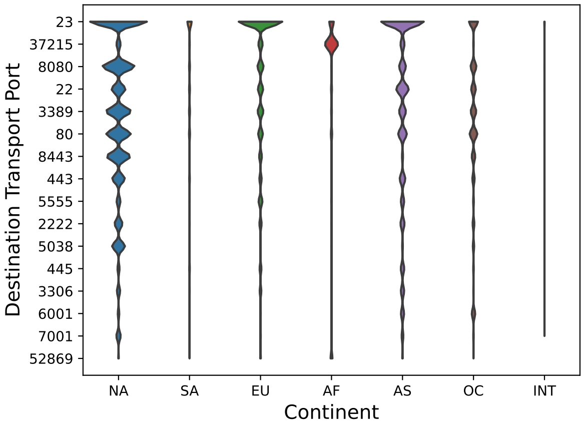


Figure 18: Bean plot of activity for the top 16 destination ports in meta-telescope traffic per world region relative to overall traffic.

D TARGETED PORTS BY NETWORK TYPE IN NA AND EU

Given the significant differences in port activities by region and by network type we now take a closer look at the two regions with the largest meta-telescope prefix space, namely NA and EU. Hereby, we separate the meta-telescope prefixes by network type, see Figure 19 and Figure 20.

For NA we again see that multiple ports are popular. We also see that port 80 is of particular interest in data centers and educational networks. The same is true for the database port 5038. For Enterprise and ISPs networks port 3389 stands out. Port 23 is still the most prominent port. For EU port 23 dominates by far. At the same time, similar observations with the NA hold. Interestingly, port rankings differ among NA and EU scanned prefixes. Hereby, we note that the differences in the percentages for ranks 4–8 are minimal. Still, port 37215 moves to rank 6 for EU and is no longer in the top 10 for NA. In addition, port 7001 gets introduced into the NA region.

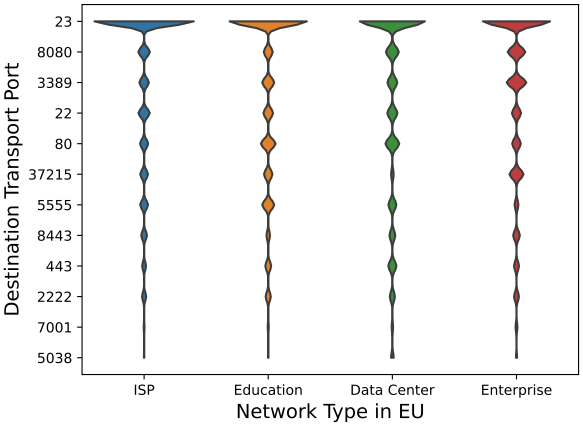


Figure 19: Bean plot of activity for the top 12 destination ports in meta-telescope traffic per network type for regions EU.

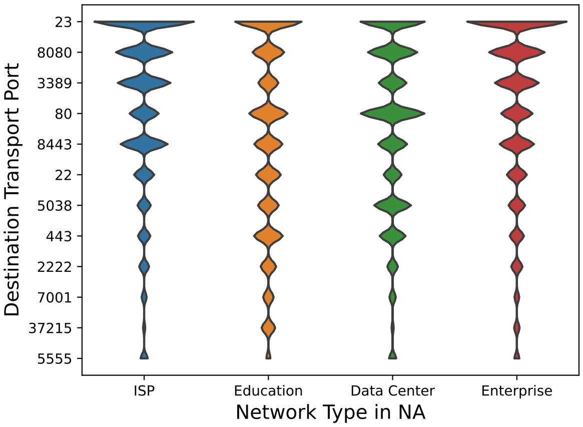


Figure 20: Bean plot of activity for the top 12 destination ports in meta-telescope traffic per network type for regions NA.