Delft University of Technology

IN4253ET: "Hacking Lab": Applied Security Analysis

Final Report

Team
Alberto Castagnaro (5861489)
Nathan Deridder (5839777)
Frank Broy (5016894)
Adrian Kirikal (4917308)
Daan Prinsze (4346106)

Supervisor
Dr. Giovane C.M. Moura

April 14, 2023

IN4253ET: Hacking Lab

ABSTRACT

The purpose of this paper is to replicate and extend the work of the "Scanning the Internet for Liveness" paper and to add insight and answer sub-questions on how the reachability of autonomous systems and subnets changes over time, across different vantage points and with different protocols. In order to do so, we collected data from different datasets, namely the ANT Lab dataset, the Censys dataset, and the CAIDA dataset. The analogies and the differences regarding several factors that we find and present, are based on this collected data, by selecting multiple fixed parameters and one free variable: time (specific short time period or evolution over the years), protocol (ICMP, TCP on port 22 and port 80, UDP on port 53), and vantage point (we analyse some of the vantage points from which the scans have been performed). Additionally, we select the autonomous system 3320 for further analysis. Finally, we try to hypothesize the possible causes of these results.

1 INTRODUCTION

Scanning the Internet can serve multiple purposes, such as checking which IP ranges are used, which services run on certain addresses, and finding vulnerabilities in those services. Moreover, one could also map out the topologies or check which protocols are supported. Scanning involves sending probe packages to the hosts and waiting for a response. If the probed host responds, then we know that it is alive and we can start narrowing down our scan, namely to check if the host rejected our request, if it accepted the request, or if an error was returned. With the help of these responses, it is possible to get a clear image of which ranges of IPs are used, which are unused or offline, and which are open to the Internet.

Next to normal pings, a scan for applications on specific ports could also be done, which could map out even further which specific applications run on a target host. This however goes into a gray area, as specifically scanning for those services is considered illegal in certain places, so this would have to be done with caution. Attackers usually scan their victims for open ports and running services, in order to find vulnerable versions of software running on their network. So in case we decide to scan certain ports, it would be the best idea to stick with the most common ones, such as HTTP (80), HTTPS (443), FTP (21), SSH (22), SMTP (25) and more. Another factor that has to be taken into account when scanning the Internet is to not harm the hosts that are being scanned. So one must be careful not to accidentally crash a network or send too much traffic and cause a Denial of Service. Some safe scanning practices are discussed in [5].

Internet scans can also be conducted over various protocols, as they may yield different results based on which protocol is used, as some may be prioritized by some hosts over others. For example, only probing over the TCP protocol may yield different results than also pinging over UDP or even ICMP. With all these measurements, the prioritisation could be measured by checking the response times across the different protocols. Also, different protocols also allow for different sorts of scans, based on how the protocols work. For example with the TCP protocol, a handshake could be established with the host, or a simple RST packet could be sent. Based on these packets, different responses could be received which can be interpreted in different ways and would give us a lot of information about the hosts themselves. Tools like NMap or ZMap [3] can be used to conduct these scans.

The Internet is a big network that is made of a lot of smaller groups of connected devices. Autonomous systems (AS) are groups of networks under a common administrative domain that operate as a single entity. These systems use the Border Gateway Protocol (BGP) to exchange routing information with other autonomous systems on the Internet. BGP tables contain information about the paths that packets can take to reach different destinations on the Internet. Autonomous systems use BGP to advertise the IP addresses they control and the routes to those addresses. This information is then propagated to other autonomous systems, allowing them to build a complete view of the Internet's topology. Autonomous systems can also use BGP to control how traffic flows through their network by manipulating the routing information they advertise to other systems. This allows them to optimize traffic flow, manage network congestion, and ensure high availability and reliability.

During our research for this paper, we did not conduct any scans ourselves, but we used the existing data from various sources to answer various different questions about the reachability of autonomous systems. We used the dataset from the ANT lab, a research group from the University of Southern California (UCS), Colorado State University, and Los Alamos National Laboratory. This dataset contains scans from the Internet done on the ICMP protocol, spanning various vantage points. Furthermore, we also used data from the Center for Applied Internet Data Analysis (CAIDA), which mainly contained various prefixes from different Autonomous Systems. Lastly, we also used the data collected by Censys [2], which contains TCP and UDP data over various vantage points. Since we did not do any scanning ourselves, we had to analyze the data that we acquired and try to draw conclusions from it.

The main goal of this paper is to replicate the work of the "Scanning the Internet for Liveness" [1] paper and to add some insight to it. The main goal of this research paper is to answer the question "How does the reachability of autonomous systems change over time and over different vantage points?". So by using the data collected by Bano et al., and conducting the same experiment in 2023, we can compare how the IPv4 range has evolved in the last 5 years. More specifically, we try to answer the following subquestions:

- How does the reachability change from one vantage point to another?
- How does the reachability change over time?
- How does the reachability change over various protocols and ports?

This report will cover related work in section 2 and talk about different research papers which dealt which scanning the internet and different tools which were developed for it. In section 3, the way we analyze the data that we acquired will be explained, and how we try to draw conclusions from it. Afterwards, the results will

be presented in section 4, and finally a conclusion will be drawn and the research questions will be answered in section 5.

2 RELATED WORK

The internet is in essence a very large collection of connections between systems. These connections are established based on addresses adhering to a set of different protocols which allow one system to connect to another. In this paper, when discussing address space, we mainly consider the IPv4 protocol, which is a collection of 2^{32} different internet protocol (IP) addresses. IP addresses are distributed largely by internet service providers (ISPs) in order to ensure that everyday users can have access to the internet. Additionally, enterprises such as web stores and search engines make use of Domain names in order to substitute IP addresses, as those are hard to remember. These domain name usage rights are delegated by domain name registrars, which are accredited by the internet corporation for assigned names and numbers (ICANN), which is a corporation charged with overseeing this distribution.

Because of the wide range of available applications and services on the internet and on user computers, sockets and port numbers have been created in the transport layer (layer 4) of the Open Systems Interconnection (OSI) model to allow for processes to function and direct data to those specific services, secluded from the rest. Many of these port numbers are reserved in order to provide a channel for standardised protocols such as FTP, SSH, SMTP, and HTTP [7].

In order to send data or establish a connection, there are a variety of different protocols available, each with a different goal in mind for each layer. The transport layer utilizes two protocols that are of note, the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP), whereas the internet layer (layer 3) has just one of importance, the Internet Control Message Protocol (ICMP). The TCP protocol allows for the establishment of a connection between services and applications, and the UDP protocol provides a connectionless way of communication that is used for configuring local IP addresses by means of the dynamic host configuration protocol (DHCP). The ICMP protocol fulfills diagnostic purposes and is used mainly for debugging purposes [7].

Using ZMap, a scanning tool that utilizes multiplicative grouping to perform these scans at high speeds [3], the entire available IPv4 space can be scanned and analyzed in order to reveal information about the services that are provided on each available IP address. This information allows someone to determine how secure the configurations at these addresses are as ZMap probes for open and closed ports. Furthermore, using different vantage points from which each scan is conducted, exposes information about the prioritization of different addresses over others. Conducting scans from different vantage points also provides information about how certain areas are prioritised or blocked from sending packets to a part of the address space.

A lot of information can be deduced based on the liveness of the network as well as its resilience. For example, before ZMap was created, Schulman et al. [6] showed that network connectivity was more likely to fail in the presence of extreme weather by measuring response time using a custom tool called ThunderPing that sends ICMP echo messages to each defined IP address. When conducting the experiment they also deployed this tool from different vantage points to distinguish between congestion and bad endpoint connectivity.

3 METHODOLOGY

In this section, we go over how we answer the research question as well as the various subquestions. We have received data related to IPv4 addresses from both Censys [2] and the Analysis of Network Traffic (ANT) research group [4]. Censys is a company focused around scanning and collecting information about all publicly accessible devices, and ANT is a research group from departments of various different universities with a similar goal. The dataset from Censys contains recurrent scans of the IPv4 space, with a specific list of detailed information on the responsive services, which autonomous systems it belongs to, location, and more for every IPv4 address. The ANT project has given us many different datasets, specifically a yearly census starting from 2013 all the way to 2022, with each dataset having around 5 different vantage points. However because we are only interested in seeing the changes since the last paper on this topic, we will only use the datasets from 2018 to 2022.

In order to accomplish this we need to be aware of the variables that affect the results of these scans:

- **Time**. The time and date of a scan are highly influential on the results it produces. For example, over the span of one or more months, there is a chance that a host has either changed or shut down, which would influence the result.
- **Vantage Point**. The source from which scans are conducted also has a large influence on which IPs respond back. For example, a webshop specifically meant for the Indonesian market might not respond to a request from Europe or America.
- **Protocol**. The protocol used to determine whether an IP is available or not is also an important factor, as some hosts might accept TCP requests but might not allow for UDP connections.

To analyze the effects of these variables, we can keep two of them fixed while changing the third. In this paper we aim to analyze the three different combinations of these problems:

- Variable time, fixed vantage point and protocol in subsection 3.1
- Variable vantage point, fixed time and protocol in subsection 3.2
- Variable protocol, fixed time and vantage point in subsection 3.3

In order to obtain more precise and focused results from the datasets used in this research, a decision was made to restrict a part of the analysis to a particular Autonomous System number, namely AS3320. This AS, operated by Deutsche Telekom AG, is a major telecommunications company with a vast network infrastructure. The rationale behind selecting AS3320 for this study was due to its status as one of the largest and most interconnected AS globally.

3.1 Varying Time

In this section, we analyze how the reachability of autonomous systems has evolved over time. To do this, we fix a specific port from the Censys data as well as one vantage point from both the ANT and Censys datasets. We then compare how these datasets have evolved over the time period from February/March 2018 to March/April 2022 for ANT data and from November 2020 to June 2022 for Censys data. We choose this time range as the liveness paper [1] was published in 2018 and our ANT dataset goes until March/April of 2022. We therefore want to see what has changed in this period of time.

The ANT dataset only uses ICMP, meaning that we do not have to actively fix the protocol or port. Censys on the other hand scans many hundreds of ports. We decide to go for port 53 for UDP, and ports 22 (SSH) and 80 (HTTP) for TCP, as these are the ports that have also been used in the Liveness Paper [1]. Once we have this data, we can then compare the overlaps of the datasets from various time frames. This allows us to analyze the evolution of IP subnets over the designated time period.

3.2 Varying Vantage Point

For this next experiment, we compare the visibility of different vantage points, which can be done by choosing a specific time period as well as a protocol. Luckily both Censys and the ANT dataset provide multiple different vantage points for each scan they complete, with both the ANT data and Censys data having 5 vantage points each, which stay consistent from at least 2018 to 2022.

We then check the overlap of reachability of autonomous systems from the five different vantage points in order to determine the reachability of each IP depending on where the request comes from. Analyzing this data can give us an idea of how some connections might be altered depending on the origin of the request.

3.3 Varying Protocol

Finally, the last experiment consists of changing the protocol while keeping the vantage point and time period the same. The initial idea for this was to compare the ANT database which uses ICMP with the Censys data which has data on numerous ports, but not on ICMP. The timeframe would have been the last ANT scan we have access to, which is from 19/03/2022 to 19/04/2022, during which Censys has a scan from 05/04/2022. The issue with this approach however was that none of the ANT vantage points use the Autonomous System as the Censys vantage points, meaning that we cannot actually fix the vantage point. This means that we cannot get reliable information to make an accurate assessment of the visibility of ICMP compared to other protocols like UDP and TCP.

However since Censys collects so much data on so many different ports, we can still do a protocol comparison using just the Censys data. These would all be done at the same time and from the exact same vantage point. Specifically, we would compare the aforementioned ports, namely ports 22, 53, and 80. This means that we can compare what IP addresses run which services and protocols like UDP, TCP, SSH, and more.

3.4 Usage of IPv4 subnets

When processing the internet-wide scan results, we had to decide on a suitable unit of analysis. In this study, we opted to analyze /24 subnets rather than raw IP addresses. This decision was based on our need to balance the granularity of the data with the computational complexity of the analysis. Using raw IP addresses in our analysis would provide a higher level detail but processing up to a billion addresses for each scan proved to be excessively high. By contrast, analyzing /24 subnets offered a reasonable balance between granularity and computational feasibility.

4 RESULTS

4.1 Changes over time

This subsection aims to examine the evolution and differences in the visibility and reachability of subnets on TCP port 80 from the Censys dataset and ICMP from the ANT Lab dataset over time. To evaluate how the visibility has evolved over time, several parameters have been fixed: we selected ICMP for the ANT dataset and TCP on port 80 for the Censys dataset. Additionally, we selected fixed vantage points for both protocols: vantage point e1 (Phoenix, Arizona) for the ANT dataset and vantage points TELIA and HE for the Censys dataset (2 of the 5 internet providers that Censys relies on to perform scans).

In Figure 1, the announced subspace is compared to the actually responsive subspace in the ANT Lab dataset. This comparison is done over time, more specifically over a period of 5 years. The dataset numbers in these graphs are taken from the names which the files had, which contained the data. The numbers in order, refer to the dates, more specifically: 79 = 2018, 84 = 2019, 89 = 2020, 93 = 2021, and 98 = 2022. In these graphs, we can see that not every subnet which is announced is reachable, which is expected, as not every subnet is active and/or accessible to the public. When analyzing the specific autonomous system 3320 in Figure 2, we can see similar results. Specific data such as the number of /24 visible subnets, the number of /24 announced subnets, and the ratio of responsive subnets over announced ones can be found in Table 5 and Table 6 in Appendix A.

The Censys dataset analysis has produced different results. In terms of overall responsiveness over time, as shown in Figure 3, there is a slightly descendant trend of responsiveness for HE and a more noticeable decrease for TELIA.

When comparing the ratio of the responsive subnets with the announced subnets and how it has evolved over time, Figure 1 highlights the ICMP ratio stability over time. One possible explanation for this is that ICMP is a lightweight protocol, adopted by most hosts and often used for testing and troubleshooting purposes. In contrast, the results obtained in Figure 4 from HE and TELIA for the Censys dataset, show varying degrees of visibility and performance over time, even when using the same protocol. HE showed better overall responsiveness compared to TELIA, with a lower decrease in responsiveness over time. Although the responsiveness for HE still decreased, it is not as significant as TELIA. Additionally, there was a period of four months where the reachability of TCP port 80 dropped for TELIA but not for HE, and we assume that this was likely due to measurement errors. This evolution over time could be due to increasing cybersecurity measures that could target more

, ,

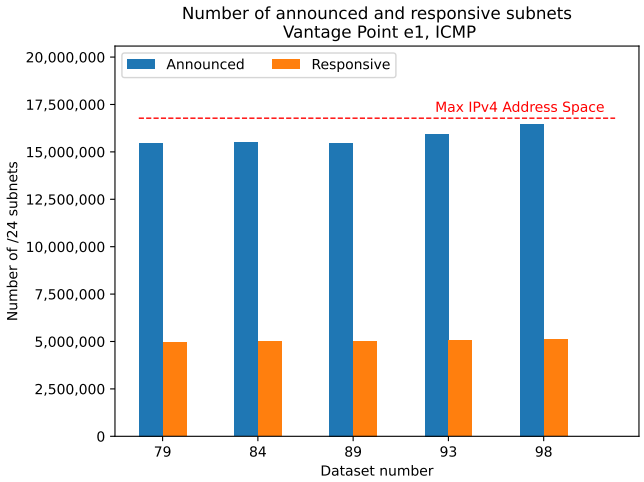

Figure 1: Evolution over time of visibility. ICMP protocol from vantage point e1, ANT dataset

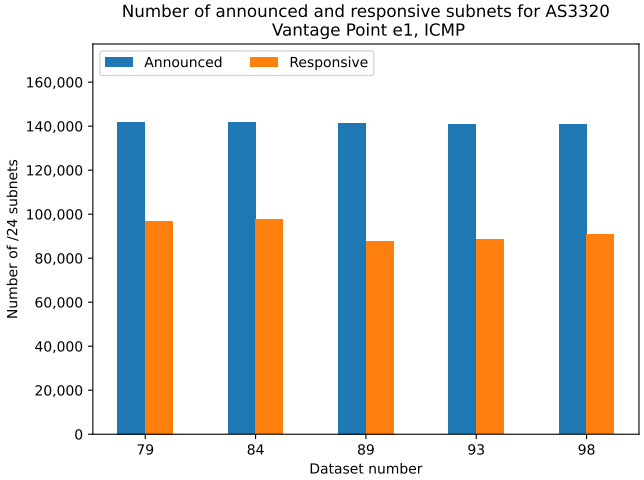

Figure 2: Evolution over time of visibility for autonomous system 3320. ICMP protocol from vantage point e1, ANT dataset

of the protocol set for the analysis or the increasing adoption of HTTPS over HTTP (as can be seen in Figure 5 and Figure 6), but there may be multiple other reasons. Specific data such as the number of /24 visible subnets, the number of /24 announced subnets, and the ratio of responsive subnets over the announced subnets for every snapshot date can be found in Table 1 and Table 2 in Appendix A.

4.2 Vantage Point visibility

In this subsection, we will analyze how visibility and reachability evolve depending on the vantage point. We conducted our research for the visibility choosing protocols ICMP, TCP on port 22 (SSH) and port 80 (HTTP), and UDP on port 53 (DNS), as well as the time

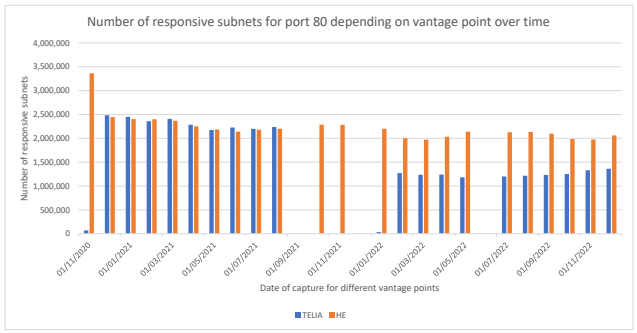

Figure 3: Evolution over time of visibility. TCP protocol on port 80 from vantage points HE and TELIA, Censys dataset

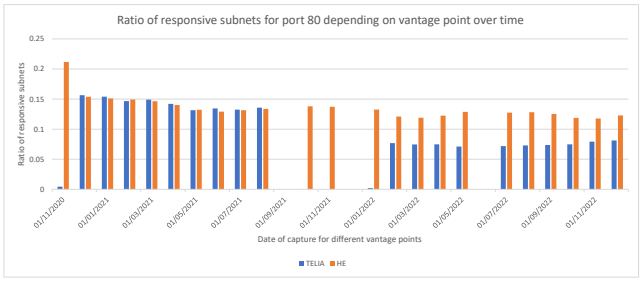

Figure 4: Ratio over time of visibility vs announced space. TCP protocol on port 80 from vantage points HE and TELIA, Censys dataset

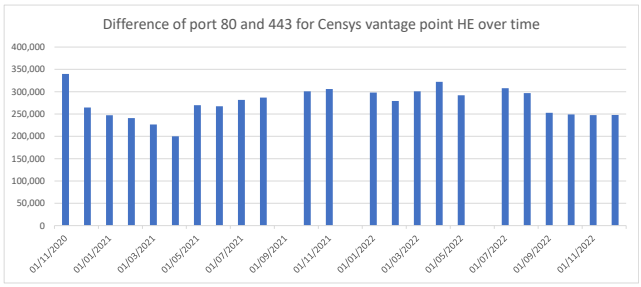

Figure 5: Difference between HTTP and HTTPS adoption over time

period, specifically March 2022. Additionally, we performed a more in-depth analysis of AS3320.

We first analyzed the reachability and responsiveness of /24 subnets over the entire IPv4 space and then for AS3320 for the ICMP protocol. The responsiveness, although considerably lower when compared to the number of announced subnets, did not vary considerably between vantage points, as shown in Figure 7. The same result was obtained in the analysis of the AS3320 visibility in Figure 8, where the responsiveness of different vantage points was almost equal.

After that, we analyze how the responsiveness of /24 subnets differs for different vantage points for TCP on port 80, TCP on port

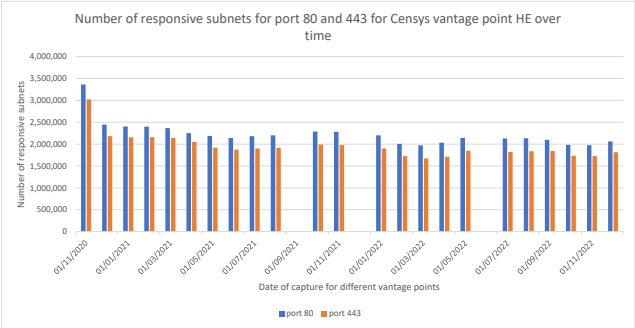

Figure 6: Comparison of the HTTP and the HTTPS protocol adoption over time

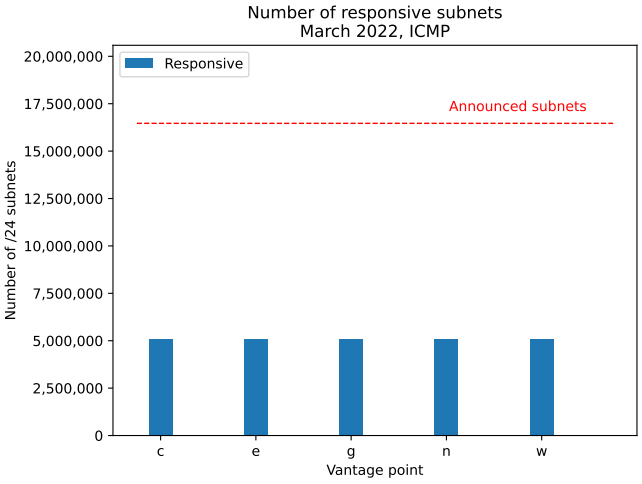

Figure 7: Comparison of responsiveness between different vantage points for the ANT dataset

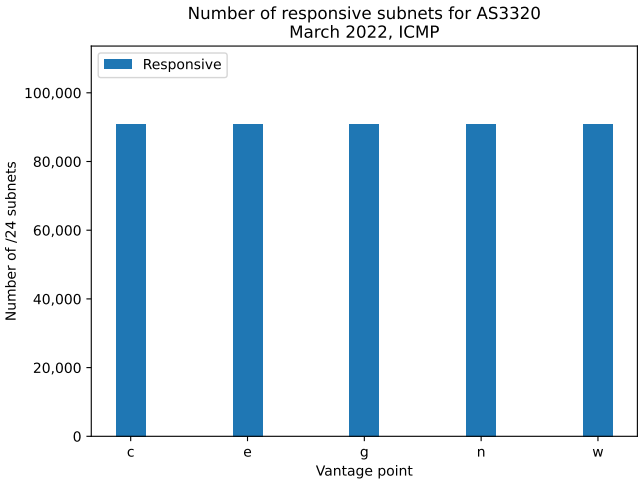

Figure 8: Comparison of responsiveness between different vantage points for the ANT dataset for the autonomous system 3320

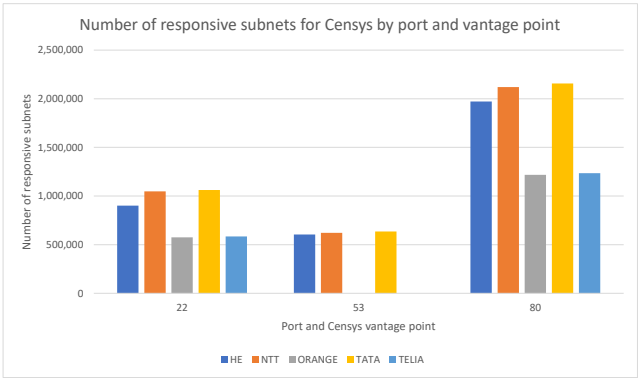

Figure 9: Comparison of responsiveness between different vantage points for the Censys dataset (March 2022), grouped by protocol

22, and UDP for port 53 over the same period of March 2022 for the Censys dataset. Here again, we performed a deeper analysis of the AS3320.

Figure 9 points out how 3 vantage points, specifically HE, NTT, and TATA, have performed better than TELIA and ORANGE for TCP port 22 and TCP port 80. This difference is even more significant when looking at UDP port 53, where TELIA and ORANGE seem to have no responsiveness at all. Additionally, NTT and TATA show slightly better reachability than HE. The analysis of AS3320 has highlighted the same differences in Figure 10. It is interesting to notice the results for the intersection of the responsive subnets from different vantage points in Table 9, where the intersections are not many. These differences in responsiveness among the vantage points can be due to multiple reasons, such as their location, and/or restrictions that may be applied to specific ISPs. However, these are pure speculations and the exact reason should be investigated more thoroughly with further analysis on a specific vantage point.

4.3 Protocol visibility

In this subsection we will inspect how responsive the subnets for AS3320 are when considering specific networking ports and their protocols. For this analysis we have taken TCP (port 22 and port 80), as well as UDP (port 53) into consideration and we will base the analysis on the collected data from five different vantage points, being HE, NTT, TATA, ORANGE and TELIA. When collecting information on the responsiveness, we have used datasets from the same month of the same year in order to prevent incorrectly comparing data that could lead to an erroneous conclusion.

When considering TCP, a quick glance at Figure 10, more specifically at port 22, shows that of all five Censys vantage points, ORANGE and TELIA are definitive outliers when it comes to responsiveness. Compared to HE, NTT and TATA, ORANGE and TELIA have a responsiveness about $\sim 62,5\%$ lower than their counterparts which is somewhat remarkable and hints at similar restrictions at

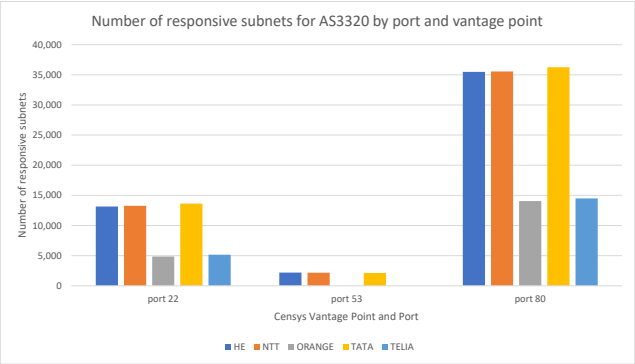

Figure 10: Comparison of the responsive subnets per port number and by vantage point

the ORANGE and TELIA vantage points when it comes to reachability.

Considering the next protocol which is UDP, which has been plotted in Figure 10, more specifically port 53, we see that both HE, NTT as well as TATA show a nearly uniform distribution when checking for how responsive the subnets are. It is important to note that TELIA and ORANGE have been omitted from this graph as Censys had not acquired any scans on port 53 from these vantage points. Another point of note is the fact that the amount of subnets that are responsive to a UDP request in AS3320 is severely lower when compared to the other protocols that have been analysed.

For the final protocol, TCP, a comparison across all different vantage points has been plotted in Figure 10 for port 80. Just as with the conducted scans on port 22, the vantage points ORANGE and TELIA once again show a responsiveness that is considerably lower compared to HE, NTT, and TATA as the responses are about $\sim 60,1\%$ less than their counterparts. However, once again it is important to note that the total number of responses received is considerably higher compared to the amount of responses received on port 22 as well as the responses received on port 53 as can be seen in Figure 10.

It is clear that the variations between HE, NTT and TATA when considering these three protocols is minor and that the same holds when comparing ORANGE and TELIA. It also becomes apparent that the rate of responsiveness when grouping HE, NTT and TATA together seems to maintain nearly the same ratio when comparing against ORANGE and TELIA when considering all three different protocols. We can only speculate as to what the exact cause is and whether these have a basis in geological position, restrictions imposed by ISPs or other factors.

5 DISCUSSION AND CONCLUSION

In this paper we have shown how the reachability of the IPv4 address space has changed over the course of the past 5 years. We have also been able to highlight the variations in responsiveness across multiple vantage points for several different protocols. We also managed to show that the difference in reachable subnets per vantage point is partially redundant, by using the datasets provided by Censys and the ANT Lab research group.

Overall, we noticed that the reachability has decreased over time for TCP, but is has slightly increased for ICMP. Furthermore, we can see that the number of responsive subnets for ICMP has stayed the same over 5 years over various ASes. By looking more closely at AS3320, we can see that this number varied over the years by approximately 10,000 subnets (around 10%).

By comparing different vantage points, we notice that for ICMP the amount of responsive subnets is approximately the same over various vantage points. This remains true, both for single autonomous systems, but also for a combination of them. For TCP and UDP, the differences are more noticeable. TCP has the biggest differences between different vantage points, but relatively they stay similar across different ports. For UDP, we noticed that some vantage points do not even have data for responsive subnets on port 53, but for the others, the numbers are very similar. Furthermore, we also detected missing values, which we suspect to be either measurement errors or other technical difficulties, which occurred while conducting the scans. By disregarding the missing values, we can still see a trend between different vantage points, which shows that the numbers are sometimes very similar (beginning of 2021), but can differ quite drastically as well (such as in 2022). As such, we can conclude that conducting scans and gathering data from multiple vantage points is crucial, as multiple vantage points help to mitigate interpretation errors as some VPs may present some bias in terms of reachability.

For the protocols, we notice that ICMP is by far the most responsive protocol, with around 5 million responsive subnets. TCP is the second most available protocol in our analysis, with around 2 million responsive subnets for port 80 and around 1 million for port 22. UDP has the lowest responsiveness, with only a few more than 500,000 responsive subnets. The reason for this is that ICMP is enabled by default in many hosts, whereas TCP and UDP need to be specifically configured. Furthermore, port 80 is one of the most used ports, which also explains the difference for TCP between the different ports that we have analysed.

So overall, scans of the Internet provide a good overview of which subnets are in use for various autonomous systems. Furthermore, by scanning the Internet, we can also compare various protocols and ports, which gives us some insights in what is running on the host. Lastly, we can also see trends in decline and increase over the years of the use of various services, such as HTTP and HTTPS for example.

6 LIMITATIONS AND FUTURE RESEARCH

One of the main limitations during this project was the budget that we received from Censys, which we used to query the data. We received a total of 300 dollars of query credits from Censys, which we could use for this project. After using all of those credits, we had to pay for additional queries and hence our queries had to be limited, as more the more data a query processes, the more it would cost. If we would have had more credits at our disposal, we could have queried more data and maybe even drawn more results and conclusions.

In the future, this research could be expanded to more autonomous systems, which could then also be compared to one another. Furthermore, more ports could be analysed and the results about different services running on those ports could be looked at in more detail

as well. This however brings up the question about legal scanning again, as those sort of scans might be considered illegal in some places.

This research could also be improved by conducting scans ourselves. The benefit of this would be that we could choose exactly which kind of data we want to collect, as well as in which time intervals we want to conduct these scans. In this scenario, we could for example choose the exact protocol, port and request, whereas by using already existing data, we do not have this freedom. We did not conduct scans ourselves due to a shortage of time that we experienced as well as due to a lack of a scanning device. We were planning on conducting the scans from a VM, where we would also setup a web page with information on how to opt-out from our scans, in case people did not want to be part of the study, but in the end, we did not get around to doing so. For future work, we would therefore look into conducting scans ourselves to improve the results.

7 ACKNOWLEDGMENTS

We would like to thank Dr. Giovane C.M. Moura for his support and valuable advice throughout the whole duration of this project. Without his active involvement in this project and his ability to provide us with hardware to conduct our experiments, this project would not have been possible.

Furthermore, we would also like to thank the ANT Lab as well as Censys, which made their data available to us upon request which enabled us to answer our research questions.

REFERENCES

[1] Shehar Bano, Philipp Richter, Mobin Javed, Srikanth Sundaresan, Zakir Durumeric, Steven J. Murdoch, Richard Mortier, and Vern Paxson. 2018. Scanning the Internet for Liveness. 48, 2 (may 2018), 2–9. https://doi.org/10.1145/3213232.3213234

[2] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *22nd ACM Conference on Computer and Communications Security*.

[3] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications.. In *USENIX Security Symposium*, Vol. 8. 47–53.

[4] ISI Analysis of Network Traffic Lab. 2023. https://ant.isi.edu/index.html https://ant.isi.edu/index.html.

[5] Vern Paxson. 2004. Strategies for sound internet measurement. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*. 263–271.

[6] Aaron Schulman and Neil Spring. 2011. Pingin'in the rain. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 19–28.

[7] David J Wetherall and Andrew S Tanenbaum. 2013. *Computer networks*. Pearson Education.

A APPENDIX

Port	Vantage Point	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
22	HE	13155	140729	0,0935
22	NTT	13258	140729	0,0942
22	ORANGE	4858	140729	0,0345
22	TATA	13631	140729	0,0969
22	TELIA	5149	140729	0,0366
53	HE	2195	140729	0,0156
53	NTT	2166	140729	0,0154
53	ORANGE	0	140729	0
53	TATA	2117	140729	0,0150
53	TELIA	0	140729	0
80	HE	35468	140729	0,2520
80	NTT	35537	140729	0,2525
80	ORANGE	14051	140729	0,0998
80	TATA	36231	140729	0,2575
80	TELIA	14480	140729	0,1029

Table 1: Number of responsive subnets, the total number of announced subnets and the ratios for AS3320 over various vantage points and ports from the Censys dataset

Vantage point	Port	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
HE	22	901336	16562240	0.0544
NTT	22	1047703	16562240	0.0633
ORANGE	22	575580	16562240	0.0348
TATA	22	1062405	16562240	0.0641
TELIA	22	583830	16562240	0.0353
HE	53	604140	16562240	0.0365
NTT	53	622291	16562240	0.0376
ORANGE	53	0	0	0
TATA	53	635714	16562240	0.0384
TELIA	53	0	0	0
HE	80	1971025	16562240	0.1190
NTT	80	2120541	16562240	0.1280
ORANGE	80	1217266	16562240	0.0735
TATA	80	2157484	16562240	0.1303
TELIA	80	1234555	16562240	0,0745

Table 2: Number of responsive subnets, the total number of announced subnets and the ratios over various vantage points and ports from the Censys dataset

Snapshot date	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
3-11-2020	69337	15889917	0,0044
1-12-2020	2491847	15887303	0,1568
5-1-2021	2456185	15903101	0,1544
2-2-2021	2364804	16081276	0,1471
2-3-2021	2409371	16141993	0,1493
6-4-2021	2291392	16071055	0,1426
4-5-2021	2179706	16540762	0,1318
1-6-2021	2231423	16568627	0,1347
6-7-2021	2200563	16585657	0,1327
3-8-2021	2239291	16474094	0,1359
7-9-2021	XXXXXXX	XXXXXXXX	XXXXXX
5-10-2021	3516	16591864	0,0002
2-11-2021	XXXXXXX	XXXXXXXX	XXXXXXX
7-12-2021	XXXXXXX	XXXXXXXX	XXXXXX
4-1-2022	33056	16577586	0,002
1-2-2022	1271905	16566630	0,0768
1-3-2022	1235859	16562240	0,0747
5-4-2022	1241152	16588700	0,0748
3-5-2022	1182128	16632822	0,0711
7-6-2022	XXXXXXX	XXXXXXXX	XXXXXX
5-7-2022	1201110	16696967	0,0719
2-8-2022	1215937	16646417	0,0730
6-9-2022	1235399	16749203	0,0738
4-10-2022	1252454	16712272	0,0749
1-11-2022	1330186	16776034	0,0793
6-12-2022	1365762	16772348	0,0815

Table 3: Number of responsive subnets, number of announced subnets as well as the ratio of the TELIA vantage point for port 80 on the TCP protocol of the Censys dataset

, ,

Snapshot date	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
3-11-2020	3375845	15889917	0,2125
1-12-2020	2452628	15887303	0,1544
5-1-2021	2407681	15903101	0,1514
2-2-2021	2404854	16081276	0,1495
2-3-2021	2367767	16141993	0,1467
6-4-2021	2258835	16071055	0,1406
4-5-2021	2190399	16540762	0,1324
1-6-2021	2143216	16568627	0,1294
6-7-2021	2183320	16585657	0,1316
3-8-2021	2204872	16474094	0,1338
7-9-2021	XXXXXXX	XXXXXXXX	XXXXXX
5-10-2021	2290279	16591864	0,1380
2-11-2021	2284674	16625566	0,1374
7-12-2021	XXXXXXX	XXXXXXXX	XXXXXX
4-1-2022	2198617	16577586	0,1326
1-2-2022	2001959	16566630	0,1208
1-3-2022	1970507	16562240	0,119
5-4-2022	2030123	16588700	0,1224
3-5-2022	2139506	16632822	0,1286
7-6-2022	XXXXXXX	XXXXXXXX	XXXXXX
5-7-2022	2125091	16696967	0,1273
2-8-2022	2131322	16646417	0,1280
6-9-2022	2098967	16749203	0,1253
4-10-2022	1986224	16712272	0,1188
1-11-2022	1975995	16776034	0,1178
6-12-2022	2062798	16772348	0,123

Table 4: Number of responsive subnets, number of announced subnets as well as the ratio of the HE vantage point for port 80 on the TCP protocol of the Censys dataset

Dataset/Year	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
79/2018	4936809	15469325	0.3191
84/2019	4995517	15489712	0.3225
89/2020	5034423	15467334	0.3255
93/2021	5056208	15951475	0.3170
98/2022	5113786	16464748	0.3106

Table 5: Number of responsive subnets, number of announced subnets and the ratio per year from the ANT Lab dataset from vantage point e1 (Phoenix, Arizona, USA)

Dataset/Year	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
79/2018	96859	141870	0.6827
84/2019	97687	141654	0.6896
89/2020	87914	141330	0.6220
93/2021	88403	140727	0.6282
98/2022	90862	140708	0.6457

Table 6: Number of responsive subnets, number of announced subnets and the ratio per year from AS3320 from the ANT Lab dataset from vantage point e1 (Phoenix, Arizona, USA)

Vantage point	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
c1	5109179	16464748	0.3103
e1	5113786	16464748	0.3106
g1	5108247	16464748	0.3103
n1	5109113	16464748	0.3103
w1	5115077	16464748	0.3107

Table 7: Number of responsive subnets, number of announced subnets and the ratio across different vantage points in the year 2022 from the ANT Lab dataset

Vantage point/Year	Number of responsive subnets	Number of announced subnets	Ratio (responsive/announced)
c1	90875	140708	0.6458
e1	90862	140708	0.6457
g1	90859	140708	0.6457
n1	90887	140708	0.6459
w1	90870	140708	0.6458

Table 8: Number of responsive subnets, number of announced subnets and the ratio across different vantage points for AS3320 in the year 2022 from the ANT Lab dataset

VP1	**VP2**	**VP3**	**VP4**	**VP5**	**Intersection**	**Union**	**Ratio (Intersection/Union)**
TATA	HE				512151	1433469	0,3573
TATA	ORANGE				435070	1189784	0,3657
TATA	NTT				668334	1424445	0,4692
TATA	TELIA				439145	1193855	0,3678
HE	ORANGE				312919	1150155	0,2721
HE	NTT				507486	1423513	0,3565
HE	TELIA				316342	1154878	0,2739
ORANGE	NTT				431254	1178979	0,3658
ORANGE	TELIA				334511	815943	0,4100
NTT	TELIA				435368	1183011	0,3680
TATA	HE	ORANGE			272250	1528884	0,1781
TATA	HE	NTT			397359	1694087	0,2346
TATA	HE	TELIA			274486	1531768	0,1792
TATA	ORANGE	NTT			390688	1519963	0,2570
TATA	ORANGE	TELIA			305586	1301014	0,2349
TATA	NTT	TELIA			393777	1523009	0,2586
HE	ORANGE	NTT			270577	1521071	0,1779
HE	ORANGE	TELIA			210161	1288763	0,1631
HE	NTT	TELIA			272922	1524025	0,1791
ORANGE	NTT	TELIA			303691	1292091	0,2350
TATA	HE	ORANGE	NTT		249877	1769636	0,1412
TATA	HE	ORANGE	TELIA		201218	1607201	0,1252
TATA	HE	NTT	TELIA		251732	1771985	0,1421
TATA	ORANGE	NTT	TELIA		294270	1599023	0.1840
HE	ORANGE	NTT	TELIA		200263	1600661	0.1251
TATA	HE	ORANGE	NTT	TELIA	194340	1831050	0.1061

Table 9: Number of visible subnets in the intersection and union of various vantages points