UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

# A Look into the Scanned and Non-Scanned IPv4 Address Space

Joël van der Kaaij

Thesis

September 2024

**Supervisors**
dr. R. Holz
dr.ing. F.W. Hahn

Design and Analysis of Communication Systems (DACS)
Faculty of Electrical Engineering
Mathematics & Computer Science
University of Twente
P.O Box 217
7500 AE Enschede
The Netherlands

# ABSTRACT

Internet-wide scans are common practice and usually give a lot of information about reachable hosts on the Internet. Applications for Internet-wide scanning vary from security-related to tracking progress on protocol adoption rates. Recently, these scans are executed on an almost daily basis for the entire IPv4 Internet address space. While a lot of information is gained from this for reachable IP addresses, little is known about the hosts that do not respond or a unreachable from the scanning server. Our research finds that, using the scan data from Censys, that at least one IP address was scanned from 96% of the assigned ASs (Autonomous Systems) and around 69% of the assigned BGP (Border Gateway Protocol) prefixes. Furthermore the most common usage type for BGP prefixes that are scanned are from ISPs or Data Centers and the most common usage type for non-scanned BGP prefixes is also from ISPs, followed by Mobile Internet Service Providers. When we look at where the most scanned and non-scanned prefixes are located, we can see that the United States is at the top, followed by Brazil. In percentage, 84% of BGP prefixes belonging to libraries are scanned, while only 44% of the BGP prefixes belonging to the military is scanned.

# CONTENTS

# 1 INTRODUCTION

With the still ongoing rapid growth of digital infrastructure, the need for understanding how the Internet's landscape is connected stays a relevant topic. A way to understand the scale of the Internet and the number of connected devices is through Internet-wide scanning. Internet-wide scanning is the practice of going through each available IPv4 (or IPv6 addresses) and see what services are available on each port.

Through this research we hope to give a look into the 'dark corners' of the Internet that are not reachable or not responding to scanning servers. We will classify the networks by BGP prefix on usage type and apply GeoIP databases to both scanned and non-scanned parts of the Internet. This research will take a deeper look at the address space that is not covered by other scans. We will compare the scanned internet address ranges with the assigned ranges according the the global routing tables.

## 1.1 Report Structure

The report is structured as follows; Section 2 explains the background on fundamental concepts used to understand the report. Section 3 proposes the research questions. Section 4 dives into the literature around Internet-wide scanning. Section 5 goes over the methodology and data analytics process. Section 6 displays the results following from the methodology. Section 7 summarises the conclusion of the results. Section 8 and 9 discusses the results and proposes future work.

# 2 BACKGROUND

For this research we define a couple of concepts to distinguish between scanned and non-scanned IP addresses.

**Scanned IP address / AS / BGP prefix**
A scanned IP address is a address that responds to a scanning probe. In the context of this research it means that is included the dataset provided by Censys.

**Non-scanned IP address / AS / BGP prefix**
A non-scanned IP address is a address is not reachable, or does not respond to a scanning probe. It is hard to distinguish between these cases as a not reachable IP address can appear be a non-response IP address on purpose.

**Coverage**
We define the coverage as ratio of scanned and non-scanned IP addresses, ASs and BGP prefixes.

**Reachable host**
A IP address is reachable if a connection is made between the scanning host and the host that is being scanned.

**Assigned host**
To be able to connect an IP address, a route over the Internet from the scanning host to the target should exist. A assigned and routable IP address should appear in Route-Views assigned IP address lists.

# 3 RESEARCH PROPOSAL

## 3.1 Problem Statement

While Internet-wide scans uncover the parts of the Internet that can be reached, it does not uncover the parts that cannot be reached or do not respond to scanning probes. Most of the Internet-wide scans conducted focus on the reachable parts of the Internet. These studies look at the information they can get from reachable hosts and extract lists of services provided for each host. This is exactly what companies like Censys do. Less investigated is the parts that are not scanned. While it is interesting to know what devices expose to the public Internet, it would also be interesting to see what is known of the non-scanned parts of the Internet.

## 3.2 Research Questions

To get insight in these undiscovered parts of the Internet we will take a look at the IPv4 address space that is not reached in these Internet-wide scans, but is assigned according to global routing tables.
For these reasons we will answer the following three research questions:

Q1. What portion of the internet is not scanned (or is not responding to scans, or is not reachable by scanning servers)?

Q2. What is the usage type of the networks that are not included in the Internet-wide scans?

Q3. What geographic part of the world is scanned and non-scanned?

# 4 LITERATURE

## 4.1 Internet Scanning Methods and Strategies

Internet wide scanning is the practise of scanning the full internet address space for responsiveness, availability or service discovery. This practice can be traced back to at least 2001, where a scanner for detecting SSH services on the internet was built. The goal of this was to detect the version of the SSH software running on servers on the Internet [1]. This was mainly for the goal of understanding the number of vulnerable servers due to critical security flaws in certain versions. Scanning at that time was slow, both limited by bandwidth and implementation.

### 4.1.1 Internet Scanning Tools

Currently the most popular tools for performing internet wide scans include *Nmap* [2], *masscan* [3], and *zmap* [4]. *Nmap* is a tool developed in 2008 that can scan Internet address ranges and scan for open ports [5]. *Nmap* and *zmap* are both optimized for different use cases. *Nmap* is more optimised for scanning for open ports on a smaller number of hosts, while *zmap* introduces an efficient technique to scan the entire IPv4 address space on a single port. Running multiple instances of *zmap* can scale up the scanning process to scan on multiple ports.

Scanning the internet can be achieved on different levels. At the network level, hosts can be scanned using ICMP probes to check whether a host is alive and reachable. A host does not have to run active services to respond. On the transport layer, hosts can be scanned using TCP to detect if a host is active and operational [6].

*zmap* is a tool developed in 2013 that makes use of specially crafted TCP SYN packets that makes the scanning process more efficient [4]. It can perform a full scan against the entire IPv4 address space within hours depending on the available bandwidth. A full scan against all $2^{32}$ (4,294,967,296) IPv4 addresses on single port from a 1Gbit connection can be completed within one hour.

A more optimised version of zmap is published in 2014, one year after the initial release of zmap [7]. This new version provides better support for faster connections, so scans take less time to execute. This development can again increase the scale of scans across multiple ports.

### 4.1.2 How a scan works with *zmap*

*zmap* uses a generator of a multiplicative group to randomly cycle through all the possible IPv4 addresses. Using this method, it is not necessary to store a predefined list of IP addresses to go through, saving time and storage space. Using a given seed, it is always possible to reconstruct the list of IP addresses and the order of the list. It excludes the IANA reserved address spaces, since they are not in use or used for other purposes, such as multicast.

Raw packets are sent at the ethernet layer instead of the network layer, preventing the Linux kernel from handling all sorts of lookups for routing that can slow down the scan.

*zmap* sends a TCP SYN-packet. A target host will respond with a TCP SYN-ACK-packet, and the kernel will respond with a TCP RST-packet to close the connection.

When a TCP SYN-ACK packet is received, the source and destination port is checked to check whether the response is caused by the scan and not by something else.

There are three types of responses possible from a scan with *zmap*:

1. Unreachable/Does not respond: SYN sent, no answer.
2. Port Closed/Host exists: SYN sent, answered with RST.
3. Port Open/Service active on port: SYN sent, answered with SYN-ACK, RST sent (by kernel).

Using built-in target generation, it is possible to generate a random list of IP addresses to be scanned. This uses a mathematical generator so a list does not have to be written to disk and can be interrupted and continued at any time. Using a randomised list of IP addresses has the advantage of not overloading IPv4 prefixes, contrary to scanning sequentially.

### 4.1.3 Scanning Ports

While *zmap* is extremely efficient in scanning a single port across the entire IPv4 address range, it cannot be used to scan all services across all the 65335 ports that can be in use for services. A scan to discover all open SSH ports is limited to scan on port 22. An SSH service on a random other port like 12345 would not be discovered using this strategy. This would require a separate SSH port scan on port 12345.

According to a recent paper, many services do not run on their assigned port number. The authors conclude that 97% of HTTP services do not run on port 80 and 93% of TLS services do not run on port 443 [8].

Another recent paper discusses an algorithm to predict services running on ports other than their default port [9]. This paper uses knowledge of services that are usually used together. An example of this is a web server running on port 80, with a mail server on port 21, a MySQL database service on port 3306, and an SSH server running on port 22. When a web server is found to be running on port 80, it is likely that other services

are used as well on the same host, so it scans the other ports as well to discover the other services.

#### 4.1.4 Consideration on the Location of Scans

According to a study, the origin location of an internet scan can influence the results of the scan. The study reports that scanning from a single origin can miss 1.6–8.4% of HTTP, 1.5–4.6% of HTTPS, and 8.3–18.2% of SSH hosts. The conclusion is that future researchers should take the location they scan from into consideration when performing Internet-wide scans [10].

#### 4.1.5 Continuous Scanning

Companies like *Censys* [11], *Rapid7* [12], and websites such as *Shodan* [13] continuously scan the entire IPv4 address space. As scanning the whole internet takes some time, it is more a snapshot of the internet during the period of the scan. Scanning the whole IPv4 address range periodically shows valuable information about the state of the Internet at a period in time. It can show how the Internet evolves through time, appearance and disappearance of hosts, services, and protocols. It can uncover and show vulnerabilities over time and show when these are patched. A more elaborate study of applications can be found in section 4.2.

#### 4.1.6 IPv6

While the IPv4 address space is limited in number to $2^{32}$ addresses, the IPv6 address space is much larger with $2^{128}$ addresses. IPv4 addresses can easily be enumerated and be scanned by either sequentially scanning the IPv4 address range or by using efficient target generation [4]. The IPv6 address space is too large to brute-force, rendering methods used for IPv4 out of use.

One of the methods for target generation for the IPv6 address range is to use lists of known active hosts. These lists can be obtained, for example, from DNS services, CDNs (Content Delivery Networks), or by analysing sources of HTTP traffic. When a host connects to one of those services using their IPv6, it can be added to the list of known active hosts to scan. More sophisticated approaches include looking at clusters of active IPv6 addresses and generating possible targets from those [14].

As there are currently no consistent and complete methods to enumerate active IPv6 hosts, further research into IPv6 scanning is out of scope for this report. Future research in this field might provide better methods to obtain known active hosts.

### 4.2 Applications of Internet Scanning

Since the development of tools to efficiently scan the entire IPv4 address space, there have been many applications for such scans. This ranges from discovering vulnerable services on open ports on hosts to tracking the adaptation of new protocols, such as the shift from HTTP to HTTPS.

### 4.2.1 Vulnerability Scans

Internet wide scans can be used to scan and check for known security issues on the Internet. It can, for example, help identify the number of vulnerable devices when a newly discovered vulnerability is announced. A recent example is a case where a new critical vulnerability is found in the popular OpenSSL software [15]. Using the Internet scans that Censys continuously does, it could quickly be determined how widespread the use of the vulnerable versions was at a period in time, including exact numbers and specific grouping by version and country.

A recent IMC 2022 paper of an HTTP(S) Internet wide scan showed that installer web pages of web applications are visible on the Internet [16]. Such pages, meant for initially setting up the web application, are especially vulnerable because no or default passwords are used so an application can be actually installed. This poses risks when an attacker actively searches for those pages to hijack the installation process and compromise the web application. In addition, unintentionally installed applications can remain vulnerable since the owners of a host might not be aware of its existence.

A 2016 study shows that by scanning several protocols used by Industrial Control Systems (ICS) a total of around 60.000 systems were publicly exposed to the Internet [17]. This is a risk since often these types of systems control large industrial machines or physical infrastructure. Since these industrial specific protocols are designed to be used in a closed network that is disconnected from the internet, it is an important find that so many devices were still exposed to the Internet.

An earlier paper from 2015 already warned about the large number of possibly vulnerable Internet of Things (IoT) devices [18]. In general, IoT devices never get updates, not even security updates. This makes a lot of devices vulnerable when new vulnerabilities are found, and never get patched. The paper at that time showed using *masscan* and *nmap* that it could find devices vulnerable to Heartbleed on their campus network. In addition to that, they showed a proof of concept for network connected printers.

A study in 2019 used Censys' Internet-wide scanning data [11] and analysed it to check for known vulnerabilities [19]. They checked the data against Common Vulnerability databases to list services that are vulnerable according to these databases. They showed that it is possible to identify services that are vulnerable by using their banner, an string that a service return when discovered, or connected to. It often includes a version number of the software that is running on a certain port. These strings can be matched with vulnerability databases that show vulnerabilities per software and per software version.

### 4.2.2 Protocol Adaptation

Internet scanning can be useful to monitor the progress of adapting a new protocol, or monitor the adaptations of updated version of a protocol.

In 2013, HTTPS was not as widely implemented as it is today. The authors of zmap performed 158 Internet scans over a period of one year, seeing the adoption rate of HTTPS 'in the wild' [4]. Research like this is important because it shows how much

traffic on the Internet is encrypted instead of being sent in plaintext. Making a general improvement in security by encrypting passwords and sessions while being sent over the Internet.

A study on the adaptation of TLS 1.3 uses internet scanning to track the adaptation of a new version of TLS over time. Although not scanning the entire IPv4 address space, this study uses Internet scanning tools to scan a large number of domains [20].

A 2014 study looked at the state of SSH on the Internet [21]. Using a Internet wide scan it was possible to see the deployed versions of SSH. It includes a study over time to see adaptation and changes in the used ciphers of the SSH servers. This study was the largest at the time, executed before tools such as *zmap*.

#### 4.2.3 Other Applications

Other applications mentioned by the authors from zmap are for example, detecting outages from power outages, natural disasters and wars. Censys occasionally publishes data related to research performed on Internet wide scan data. For example when hurricane Fiona hit Puerto Rico, Censys was able to detect connectivity loss in that region [22]. Data like this can show the impact of events on Internet connectivity and possibly show other impacted areas like the extend of energy grid outages.

### 4.3 Blocklists and Abuse Reports

Blocklists (previously also known as blacklists), can be used in different ways. In the context of Internet-wide scanning, this is a list of networks that should not be scanned. Implementing such a method to be excluded from Internet-wide scans was already proposed by the authors of *zmap* in 2013 [4]. This list consists of IP prefixes that should not be scanned.

From a security perspective, blocklists can be used to block incoming traffic from untrusted networks, or known bad actors based on their IP address or domains. A research performed in 2021 shows that there are many open source blocklists available, but that it is very hard to compare these [23]. The blocklists often lack descriptions on how they are constructed, what data is used to construct them, and what the policy is to keep them up to date. Correctly labelling blocklist entries and comparing them between different blocklist providers is an even greater issue that is difficult to solve.

A fundamental difference in the use of blocklists is that blocklists used for security purposes are about incoming traffic, and blocklists used for Internet wide scanning is about outgoing traffic.

#### 4.3.1 IP Addresses, ASNs and WHOIS

IP address ranges are assigned to an ASs (Autonomous Systems). An organisation can ask to be assigned an ASN (Autonomous System Number) used for routing traffic for their IP address range to their routers. This list is governed by IANA [24]. This list

contains information about the organisation to whom the IP address range is assigned to.

Historical data for autonomous systems and their assigned IP address prefixes can be obtained through the RouteViews website [25]. For more information see the 2 section.

WHOIS database lookups can show the owner of an certain IP range. It contains relevant information about the owner, such as the name and organisation it belongs to, and often shows an email address to report malicious behaviour coming from the address ranges of the owner.

As an example, we can look at the University of Twente. It has its own AS and thus an ASN. The assigned ASN is *AS1133*, and its name is *UTWENTE-AS* [26]. All of this information is publicly available and also contains information to report abuse coming from the assigned IP range. For the University of Twente, this information is *abuse@utwente.nl* for spam related abuse, and *security@utwente.nl* for anything security related.

ASs can be categorized into sectors in which they are used. For example, into categories like hosting providers, internet service providers, and businesses. A study by [27] shows that using machine learning and combining data from different source, a classification can be made. ASdb can categorise ASs into 17 categories and 95 subcategories 93% and 75% accuracy. This is useful for new ASs that are registered, which happens on a daily basis.

Instead of using a database that provides organisation names for AS numbers, it is also possible to partly obtain this information from hostnames [28]. This study focusses on training a machine learning algorithm to get AS information from hostnames, by trying to create and match AS numbers in hostnames. This is not flawless, but might provide some direction to who is the owner of a router, and could complete other AS related information.

### 4.3.2 Opting out

Opting out of internet scans is currently not standardised or automated. Most research projects encourage target hosts to send an email to the scanning projects to exclude them from future scans. This is often a manual process.

Rapid7 requires sending an email to opt out of future scans [29]. This requires action from both the network operator as well as Rapid7 that should add the complaining network operator to their blocklist.

Opting out of Censys' scans requires a network operator's firewall to drop connections from Censys' IP address ranges [30]. Staying excluded from Censys's Internet wide scans requires network operators to keep up with Censys' IP ranges. Opting-out of Censys' HTTP-based scans is done by blocking a specific browser User-Agent string.

Blocklists are also not shared between research projects or scanning parties. This is probably due to privacy concerns and regulations that prohibit sharing such information.

Publicly sharing this information could also cause IP ranges on the list to become targets, since there must be some reason to be excluded, making them more interesting to scan. The different approaches taken by Rapid7 and Censys exemplify the lack of standardisation in opting out. This requires targets and organisations to act upon each scan attempt at their address space.

Reasons for being excluded from the scans are not known or not specified. So far, there have been no studies on the exact reasons why organisations want to be on Internet scan blocklists.

### 4.3.3 Outdated Blocklists / Expiration

IP addresses included on a blocklist can be outdated. The BGP prefix for a address space might have been transferred to other ASs. Addresses can be reused for other purposes than the reason they were included in a blocklist [31]. This implies that IP addresses included in a blocklist must be re-validated once in a while. The paper also discusses methods for detecting that an IP address is reused for other purposes. Reused addresses, now used for other purposes, could be excluded and make future scans less complete. Dynamic IP addresses, often used by Internet Service Provider customers, can even change on a daily basis.

The same behaviour occurs in the security field, where IP addresses included in blocklist should be reevaluated with a blocklist update strategy to prevent harmless IP addresses from being blocked [23].

### 4.3.4 Inferring Use of Blocklists

According to a study from 2021, it is possible to infer the use of blocklist used by networks [32]. Using a technique using the IPID and spoofing the source address, it is possible to detect whether a network blocks the incoming packets. Using this method it is possible to compare blocklists against networks by checking the IP addresses that are on that blocklist.

### 4.3.5 Unused Address Space

Not the full IPv4 address space is allocated to use by ASs. There are multiple IP prefixes that are reserved for specific purposes by IANA, the Internet Assigned Numbers Authority [33]. Commonly known reserved IP prefixes include 127.0.0.0/8 for local use, 10.0.0.0/8 for private or intranet use, and 192.168.0.0/16 commonly used for home routers. A large IP prefix 240.0.0.0/4, is reserved for future use. The list of reserved IP prefixes can be put on blocklists by default, since they are not supposed to be in use. Any unexpected traffic from these type of IP addresses is suspicious.

Other nonreserved and nonassigned IP prefixes can also be excluded from scanning, since they are not supposed to be in use. Although scanning these and other IANA

reserved addresses might be interesting since it is possible to advertise IP ranges, without being assigned them.

## 4.4 Automatic Detection and Blocking

Internet wide scans causes network traffic, and is not always desirable for network operators, and could even be a threat. The increased practicality of scans and the number of scans increase. Already in November 2018, in a single month, 2.2 million scans were detected [34]. With everyone able to do their own scans, both for good and bad causes, the question rises whether we would and should be able to detect them, and/or eventually block them. A complete network and port scan is often a preparation for further intrusion into a network using the information gathered about the services that are running on servers within that network [35].

### 4.4.1 Detecting Internet Wide Scans

A study in 2018 shows that network scanning activity can be measured and that patterns can be observed by analysing over time [36]. The was done with connection log data from routers and firewalls used at the Korean KAIST university campus network. The study observed a high increase in Telnet scans, which was likely introduced by an increase in IoT botnet devices. A rise in Telnet scanning activity was observed at the time that the Mirai botnet rapidly expanded.

A study from 2019 used network telescoping for detecting scans on the Internet [34]. When analyzing the network captured at firewalls, this research was able to identify internet-wide, partial internet-wide and localized scans. Data is used from 89.000 CDN hosts using 2 IPs each, resulting in a total if 178.000 IPs. The hosts are geographically distributed over the world. With this much data, patterns can be detected by combining all the data from all the hosts. When 150.000 of the 178.000 were contacted by the same source IP, within a certain time frame, the behaviour from the source IP can be identified as an Internet-wide scan.

Another study using network telescoping and honeypots to detect scans also concluded that scans can be dangerous, as they expose potential targets for exploitation. This is the case with ICSs connected to the internet [17]. This research also concluded by analysing their own honeypots that other actors are actively scanning for these ICS specific protocols. The intentions for these scans are unknown, but some of the scan sources include, for example, Rapid7 [12] and Shodan [13].

From the security perspective of Internet wide scanning, a study from 2015 shows that a scanning period can be followed by an attempt to exploit scanned devices [35]. This research suggests that a scan can be followed by a attack on a network, possibly infecting devices on that network with malware or other malicious software. It concludes that a network scan is dangerous. Not the scanning itself is dangerous, but the knowledge gained from the scan is dangerous.

### 4.4.2 Intrusion Detection Systems / SIEM

Another research shows that it is possible to detect port scans and identify them automatically [37]. Using machine learning, it is possible to learn how an attack looks. The method presented by this research can also categorize scans into different types of scans. Validation shows a high change of successfully detecting an attack. This might be useful for detecting earlier unseen scan types or patterns.

## 4.5 Best Practises and Ethics

When comparing to the real world, an Internet scan is like knocking on a house's door and noting down whether there is a response from the house owner. Internet-wide scan is a bit more tricky and would be like sending a large group of people to knock on all the doors in a whole neighbourhood. Doing an Internet wide scan on a range of ports would compare to knocking on all doors of a house, on all houses in a neighbourhood. The question rises whether this would considered acceptable behavior or downright illegal. As there is not definitive set of rules set for Internet wide scanning in practise, we can take a look at two lists of recommendations from *zmap* and the University of Twente, and applying them to a guideline set for research in the field of Information and communications.

The impact of scans also varies widely depending on the protocol or service scanned. A single ICMP (Internet Control Message Protocol) ping request is less intensive than a full HTTP GET request to port 80 or 443 to retrieve a page or just perform a TLS handshake to check certificates.

### Recommendations from the creators of *zmap*

As it is not possible to ask permission for all networks, one should be careful not to overload networks with too much traffic. The authors of *zmap* have constructed a list of recommendations for internet scans (quoted from their paper) [4]:

1. Coordinate closely with local network admins tor educe risks and handle inquiries.
2. Verify that scans will not overwhelm the local network or upstream provider.
3. Signal the benign nature of the scans in web pages and DNS entries of the source addresses.
4. Clearly explain the purpose and scope of the scans in all communications.
5. Provide a simple means of opting out, and honor requests promptly.
6. Conduct scans no larger or more frequent than is necessary for research objectives.
7. Spread scan traffic over time or source addresses when feasible.

### 4.5.1 Recommendations from the University of Twente

As an Internet scan can cause blocking of an IP address or damage the public reputation of IP ranges or the organisation that owns the IP range, the scan should be executed with care. The Internet Service Provider of the University of Twente, SURFnet, has established guidelines for performing Internet scans from the university network [38]. SURFnet's security department, SURFcert, has requirements for hosts performing scans to be transparent about reasons for scanning. One of these requirements is to give hosts a reverse DNS PTR record, so that a target host can lookup the IP address. Another one is to host a website on the host that is scanning the internet to inform target hosts about the purposes of the research.

SURFcert requires the following for performing internet scans from an university IP address:

1. Think about the impact scans can have on the target hosts.
2. Do no scan more than necessary for answering research questions.
3. Randomise the scan over the entire address space, so IP ranges and networks are not scanned at once.
4. If possible, make yourself known using messages in protocols. Such as the User-Agent field of an HTTP header.
5. Provide options to opt-out of scans.
6. Announce the scan to incident (local) response teams, so they are aware when target hosts sends complaints.

This information is relevant for a possible scan executed from the university's network in future work. The list of requirements from SURFcert closely resembles the list by the authors of *zmap*[4].

### 4.5.2 Guidelines for Information and Communications Research

The Menlo Report [39] presents guidelines for conducting research in the fields of information and communication, it is a successor of the Belmont Report. The four basic principles of the Menlo Report are:

1. Respect for Persons, research is conducted with informed consent from the targets in questions
2. Beneficence, maximize benefits and minimize harm
3. Justice, all actions involved should be handled equally
4. Respect for Law and Public Interest, be transparent in methods and results.

Applying the basic principles from the Menlo report, to the guidelines set by the authors of *zmap* and SURFnet.

Informed consent is impossible to apply for an Internet wide scan. It is impossible to ask for permission to scan for every IP address owner. Internet wide scanning does not interact with persons, and involves only machine to machine interaction. Generally in this context, an Internet-wide scan is seen as being in the public interest.

According to the recommendations of the authors of *zmap* and SURFnet, randomising the scanning space is a form of minimising harm, causing less traffic to hit the target at the same time. It should always be assessed to minimise the impact on hosts.

When providing options to opt-out in an Internet wide scan, all requests should be handled and handled equally.

Transparency is implemented by being open about how the data is collected and what happens after it has been collected. Research should be announced and traces leading back to the research should give information about the research.

# 5 METHODOLOGY

In order to discover networks that are not covered by scans, we will first have to know what addresses can be scanned, we will make use of the Universal Internet Dataset that is available from Censys [11]. We will then use this data to compare with the complete IPv4 address space that is assigned with data from RouteViews [25]. To get additional information about the non-scanned address space, we will take a look at the information available in a geolocation database for IP addresses provided by Geo2IPLocation. Challenges will include the enormous amount of data involved, and the time it requires to process the data. This requires special attention to data collection and processing, which also was a major influence on the research duration.

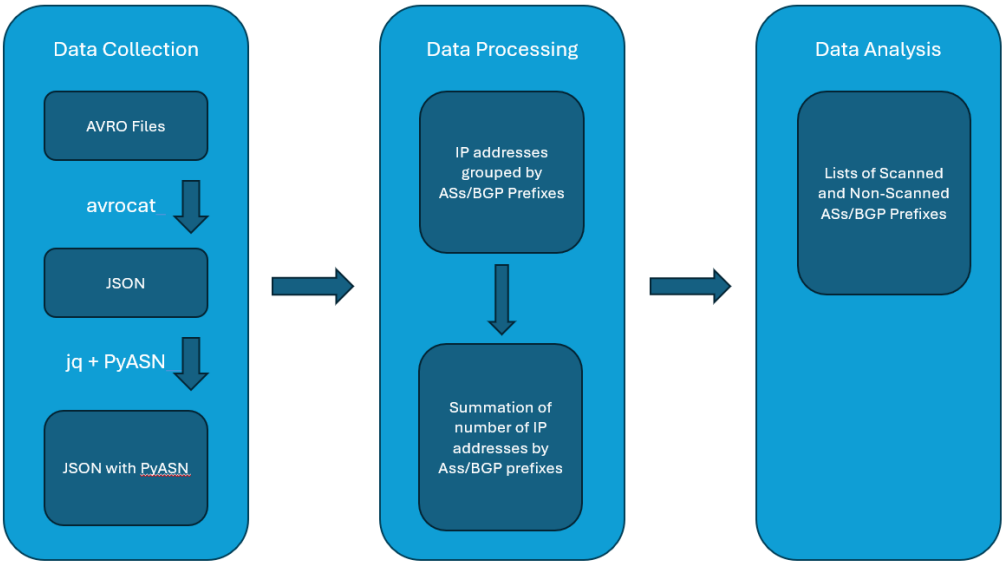


Figure 5.1: The pipeline used for data collection, processing and analysis

## 5.1 Data Collection

### 5.1.1 Internet-wide scans provided by Censys

We will make use of data provided by Censys to analyse the research questions. Censys does a weekly scan of over 3000 ports over the full IPv4 Internet address space. This includes the most commonly used ports, as well as ports that are often used by default for a range of applications. An example of this is port 80 for basic HTTP traffic, port 3306 for MySQL server, or port 5000 used by default for Python Flask projects. Data collected by Censys include the following relevant information for our analysis: IP address, port, service name, timestamp of detection.

### 5.1.2 RouteViews archive

The RouteViews Archive contains an archive of routing tables captured at different physical locations around the world. It contains a mapping of BGP prefixes to ASN and is essentially a list of assigned IP addresses. For this research, we will assume that this is the complete reachable IPv4 Internet address space.
A full dump of the Internet routing table is available for download from RouteViews every two hours.

### 5.1.3 Geo2IPLocation database

To get an idea of the usage type of unscanned networks and the parts of the world that are less represented in the scans provided by Censys, we will use the GeoIP database provided by Geo2IPLocation. This database provides a mapping from IP address range to physical location and also includes the usage type for each address range.

## 5.2 Data Processing

### 5.2.1 Censys Universal Internet Dataset AVRO files

The data provided by Censys is in AVRO format. Censys provides weekly snapshots of their scanning data, limited to a time frame of one week. The data from Censys is provided in a serialised data format for storing a large number of records, called AVRO. In addition to this, the snapshots are compressed using the snappy compression algorithm.

A single scan for Censys contains around 700GB of snappy-compressed AVRO files. For this research we will limit the data to a weekly snapshot between January 2022 and March 2023, consisting of 60 snapshots in total. This accumulates to a dataset of 42.7 TB of snappy-compressed AVRO files.

Completely extracted and converted to JSON, this would be around 2 TB of JSON structured data per snapshot, or around 120 TB of JSON files.

The dataset provided by Censys contains more that than needed for this research, so we process the data first to get the essential fields out of this dataset.

Using the tool *avrocat* we extract the AVRO files to JSON and using the tool called *jq* we select the fields needed for our analysis.

The fields we select for our research are:

```
extract_query = "{host_identifier, location, autonomous_system,
    ↪ services: {ports: [.services[].port], service_names: [.services
    ↪ [].service_name], observed_at: [.services[].observed_at]}}"
```

The resulting data from this query is 180 GB of JSON data per snapshot. To improve analysis performance later, we store the resulting data in parquet format. This is a column-orientated data file format that is more suitable for data processing and querying. Converting the data results in 20 GB of parquet files per snapshot.

### 5.2.2 RouteViews Archive RIB files

While Censys also provides the ASN and BGP prefix for an IP address, we choose the apply the BGP and ASN from the RouteViews RIB files. This gives us the opportunity to better compare the data between the global routing table and the IP addresses scanned by Censys.

The files obtained from RouteViews are processed by PyASN, a tool that converts the RIB files to a format that maps BGP prefixes to ASNs. We will use these files to define our 'universe' and we will compare the Censys scan data to this. We assume that this is the full global routing table, and that all prefixes in this table are actually assigned and routable.

Data from RouteViews is matched on the scan date of the Censys snapshot. For a Censys snapshot of the 4th of January 2022, we choose to match with a snapshot from RouteViews at 2022-01-04 00:00 UTC.

To complete our dataset we use for analysing, we attached the results from RouteViews to the Censys dataset. This leaves us with a final dataset with the size of 1.2 TB of snappy-compressed data in parquet format.

### 5.2.3 GeoIP2Location dataset

We match the GeoIP2Location to a Censys snapshot and RIB file using the GeoIP2Location snapshot day that is the closes to that of the Censys snapshot. GeoIP2Location snapshots are usually from one day prior to the Censys snapshot. The Censys snapshot of January 4th 2022 will be matched to the GeoIP2Location date of January 3th 2022.
Data analytics

# 6 RESULTS

After processing the data as mentioned in the Methodology section, we can now present the results in this section.

## 6.1 IP Address Coverage

To know which IP addresses are not scanned, we first need to know what IP addresses were present in a scan, so we can subtract these from the list of assigned IP addresses. Therefore, we analysed the weekly snapshots from Censys from January 2022 to March 2023. On average, this resulted in around 220 million unique IP addresses per scan as can be seen in Figure 6.1. The number of IP addresses scanned is relatively stable with no major drop or increase over time, such that there are only minor difference between scans.

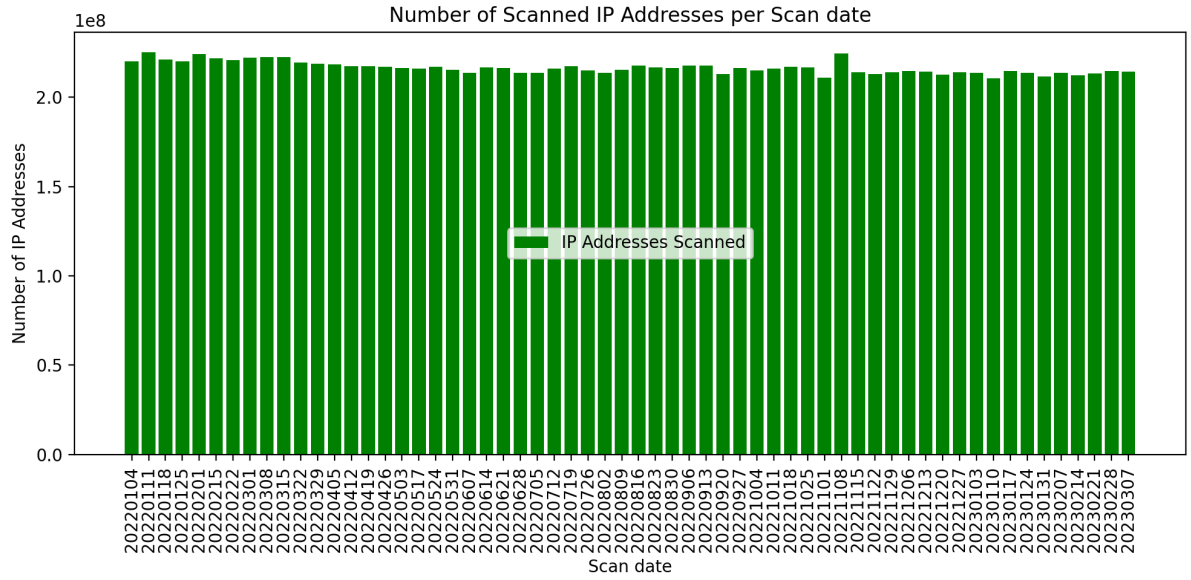


Figure 6.1: Number of scanned IP addresses per Censys scan date

Figure 6.2 compares the number of scanned IP addresses to the total number of assigned IPv4 addresses according to the data from RouteViews' RIB files. On average the total number of assigned IP address space is around 4.1 billion, where the number of IP addresses that appear in Censys' scan is between 4.97% and 5.35%. See Figure 10.1 in the Appendix for the absolute numbers.

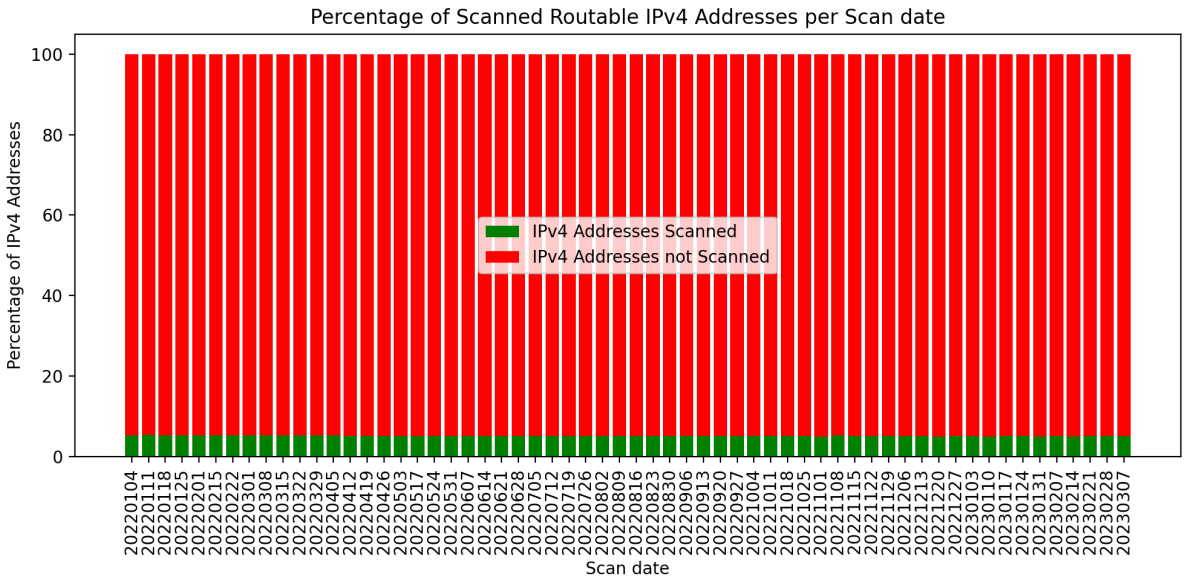


Figure 6.2: Percentage of total IP addresses and scanned IP address per Censys scan date

## 6.2 ASN Coverage

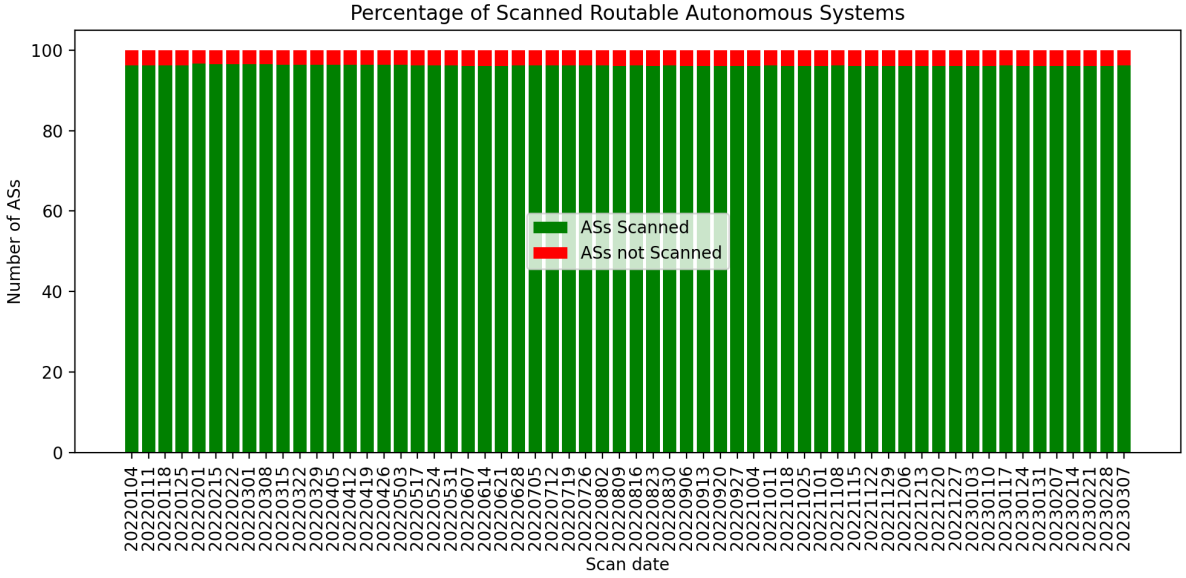


Figure 6.3: Percentage of Autonomous Systems and scanned Autonomous Systems per Censys scan date

To analyse the number of non-scanned ASs, we needed to subtract the list of scanned ASs from the list of all ASs that are active. To obtain a list of active ASs we used the data from RouteViews on the date of the scan at 00:00 UTC. For each of the Censys scan dates, we extracted the complete list of ASs from these files and compared them to the scan data that included IP addresses mapped to the same RouteViews table.

An AS is marked as scanned when at least 1 IP address was present in the scan by Censys.

As can be seen in 10.2, the percentage of scanned ASs is also relatively stable over time between 96.01% and 96.66%. However, a minor increase of the total number of ASs can be observed. See Figure 10.1 in the Appendix for the absolute numbers.

The resulting number of ASs that were never scanned over in any of the snapshots was: 1006. This number is less than one would expect, and might be because of the fact that ASs can appear and disappear over time.

### 6.3 BGP Coverage

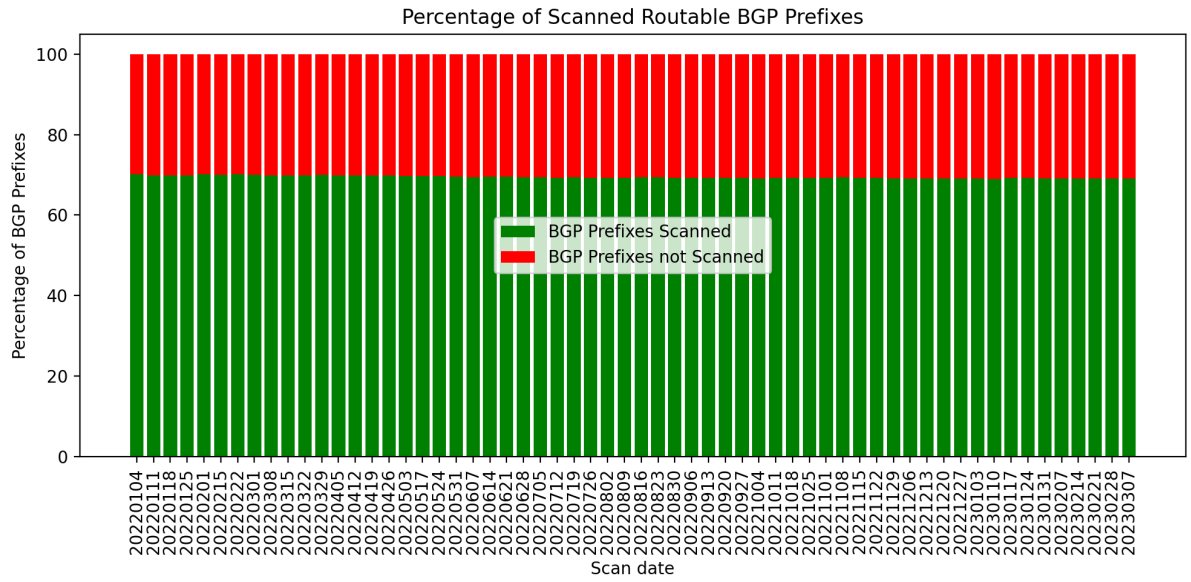


Figure 6.4: Percentage of total BGP Prefixes and scanned BGP Prefixes per Censys snapshot

For analysing the number of BGP prefixes we used the same strategy as with the ASs. For each Censys scan date, we compared the scan data with the RouteViews data to count the number of BGP prefix present in the scan.

A BGP prefix was marked as scanned when at least one IP address was present in the snapshot by Censys.

As can be seen in 10.3, the number of scanned BGP Prefixes is stable over time between 68.96% and 70.17% . A minor increase in the number of BGP prefixes can be observed. See Figure 10.1 in the Appendix for the absolute numbers.

## 6.4 Coverage in Percentage of ASN / BGP

To limit the scope and look deeper into the data of a single snapshot, we look at the Censys snapshot of January 4th 2022.

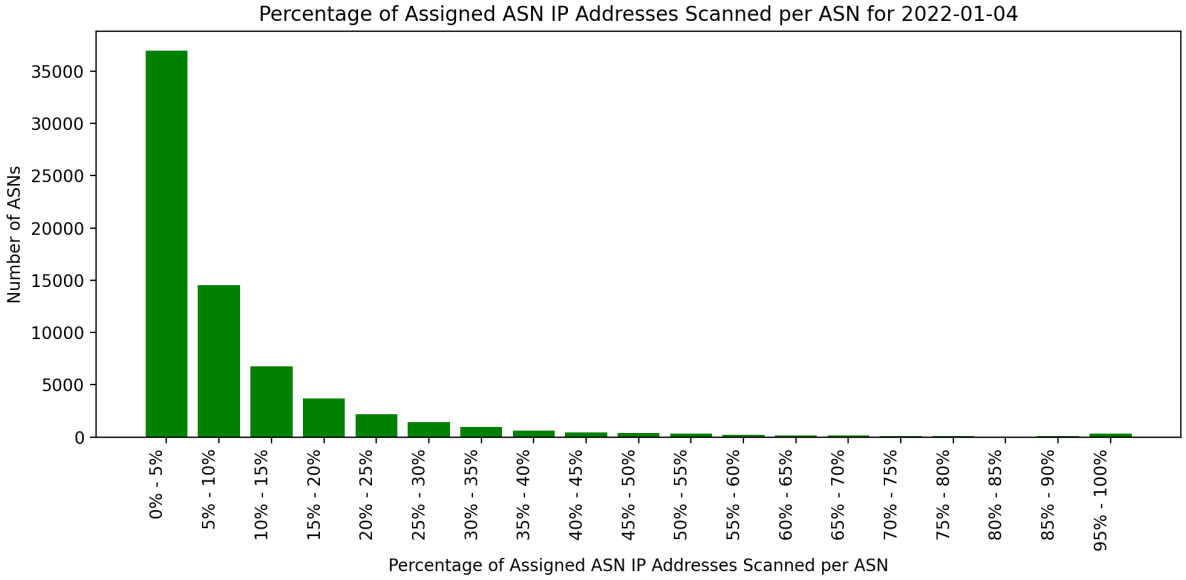


Figure 6.5: Percentage of IP addresses assigned to an AS that are scanned

To take a deeper look into the number of times a AS in represented in the scan data, we choose to calculate the percentage of scanned IP addresses, to the number of total assigned IP addresses from RouteViews. This resulted in the data represented in Figure 10.2. As can be seen, most of the ASs have a scan ratio between 0% and 15%. The full data is available in Table 10.1 in the Appendix.

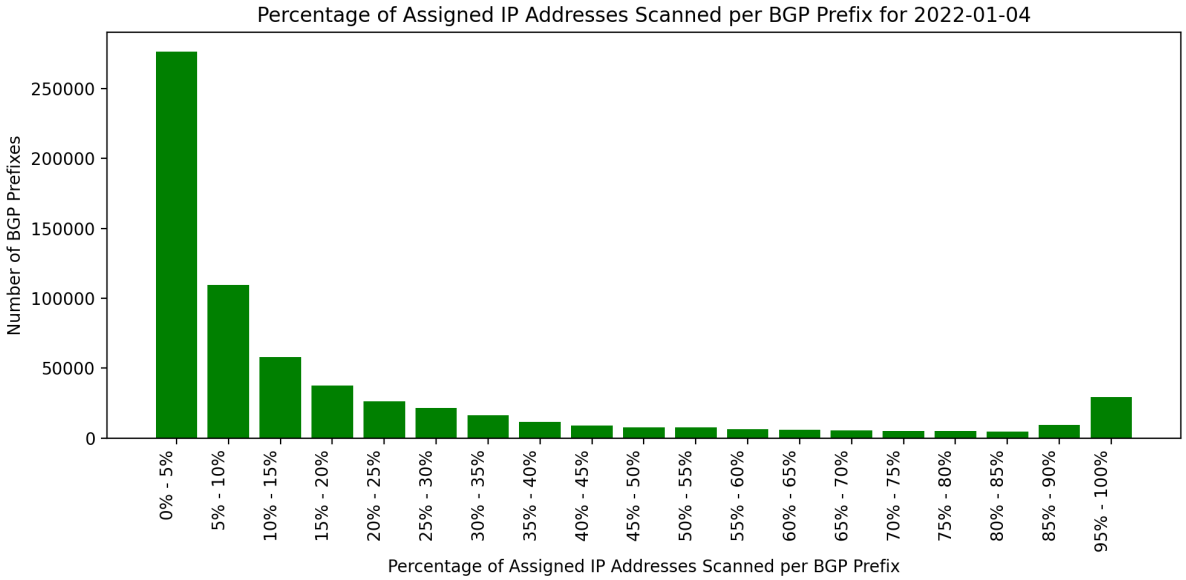


Figure 6.6: Percentage of IP addresses inside a BGP prefix that are scanned

The same approach was done for the BGP prefixes. When looking at both Figure 6.5 and Figure 6.6, we see an interesting bump at the end of the graph where 95%-100% of the ASs and BGP is scanned. This might be worth looking into in further research. A possible explanation for this is that the length of these prefixes is short, or that that there are not much IP addresses assigned to a specific AS. The full data is available in Table 10.2 in the Appendix.

## 6.5 Scanned and non-scanned BGP prefixes by usage type

The following figures show the usage type of BGP prefixes. The full results including the meaning of the abbreviations are available in Table 10.3 in the Appendix.

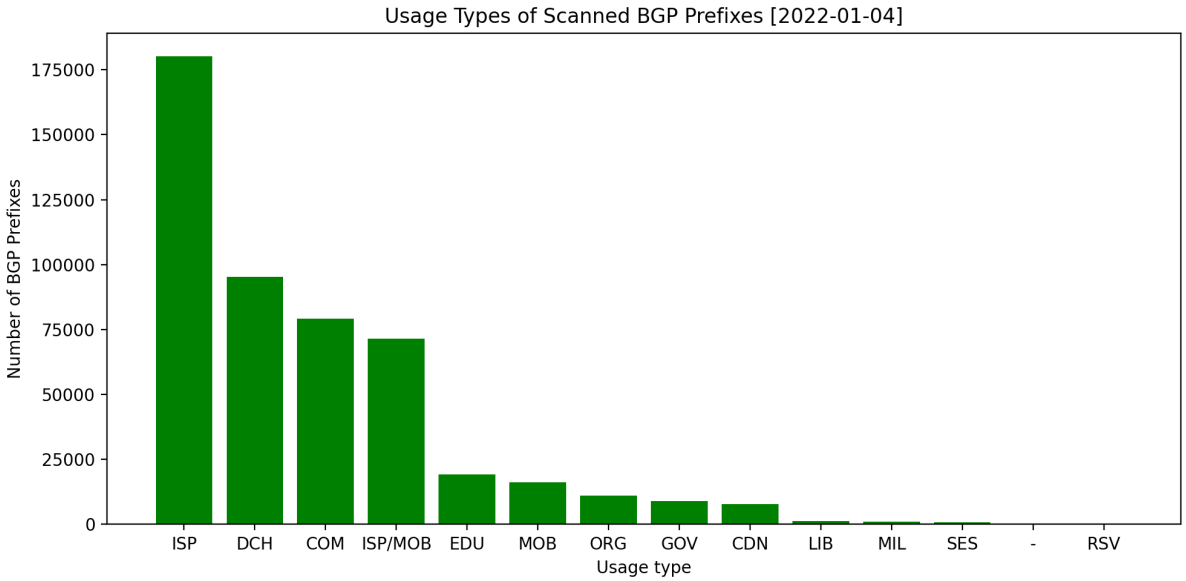


Figure 6.7: The number of scanned BGP prefixes per usage type, when the GeoIP2Location data returned multiple usage types for a BGP prefix, it is counted towards each usage type.

To see how the BGP prefixes that are scanned are used, we matched the IP address in the BGP prefixes that are scanned to the GeoIP2Location database. This resulted in a list of usage types with counts of each usage type. A BGP prefix is counted towards a usage type if the BGP prefix is within the range of IP addresses in the GeoIP2Location database.

The ISP/MOB (Internet and Mobile Service Providers) refers to ISPs that act as both fixed line and mobile operator.

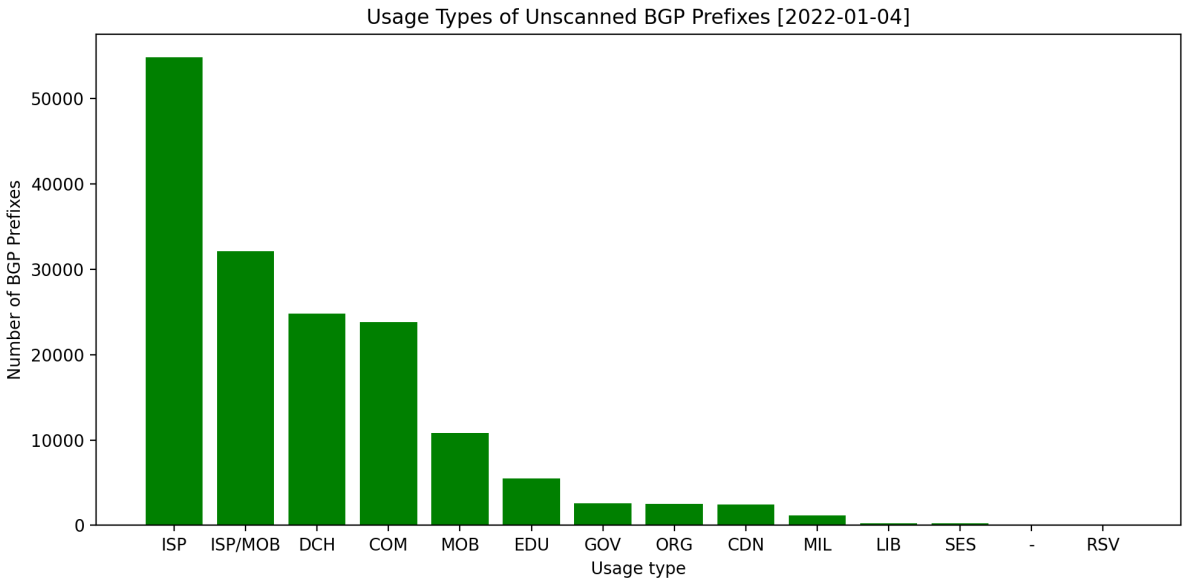


Figure 6.8: The number of non-scanned BGP prefixes per usage type, when the GeoIP2Location data returned multiple usage types for a BGP prefix, it is counted towards each usage type.

The same approach was done for non-scanned BGP prefixes. We see that the DHC (Data Center) is less dominant and that ISP/MOB is more dominant.

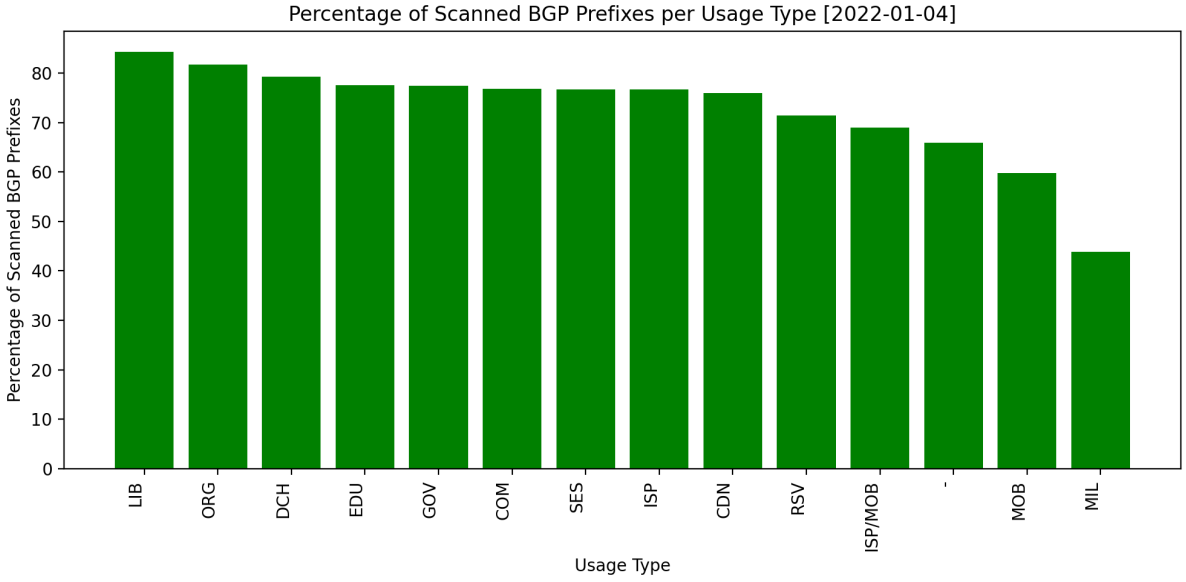


Figure 6.9: The percentage of scanned BGP prefixes per usage type, when the GeoIP2Location data returned multiple usage types for a BGP prefix, it is counted towards each usage type.

When we look in Figure 6.9 at the percentage of scanned BGP prefixes per usage type, we see that the library (LIB, 84.3%) classification has the highest percentage scanned, followed by organisations (ORG, 81.7%) and data centers (DCH, 79.3%). Less scanned are fixed/mobile operators (ISP/MOB, 69.0%) and mobile operators (MOB, 59.8%). For the BGP prefixes classified with a military (MIL) usage type, 44% is scanned.

## 6.6 Scanned and non-scanned BGP prefixes by country

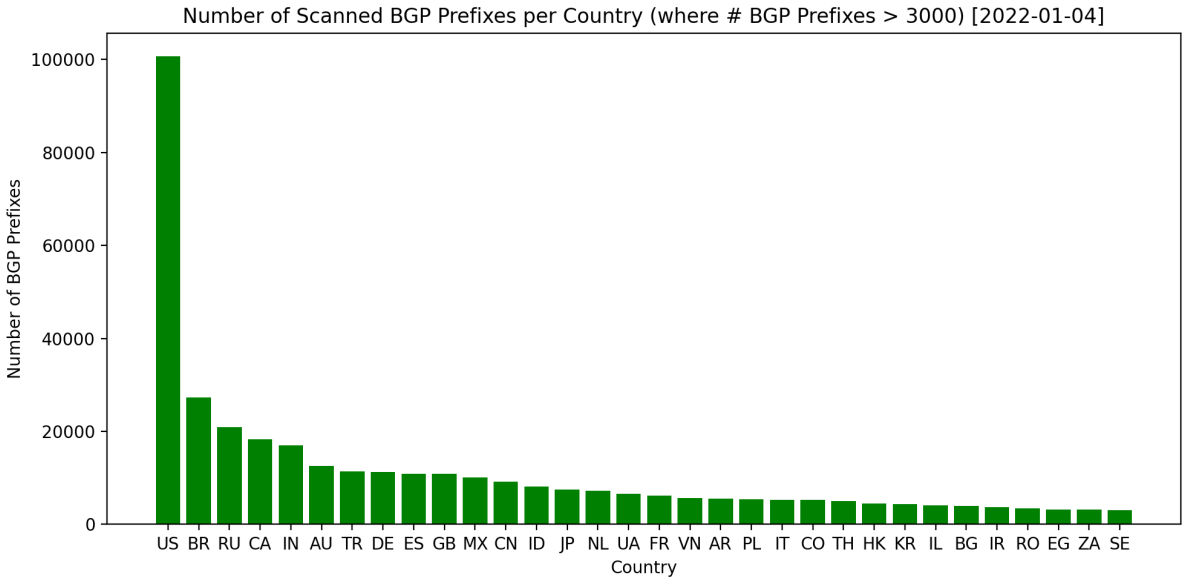


Figure 6.10: The number of scanned BGP prefixes per country, when the GeoIP2Location data returned multiple countries for a BGP prefix, it is counted towards each country.

Using the GeoIP2Location database we searched for the countries the BGP prefix is mapped to. Sometimes, multiple countries were returned for a single BGP prefix. In this case we added the BGP prefix to all the countries.

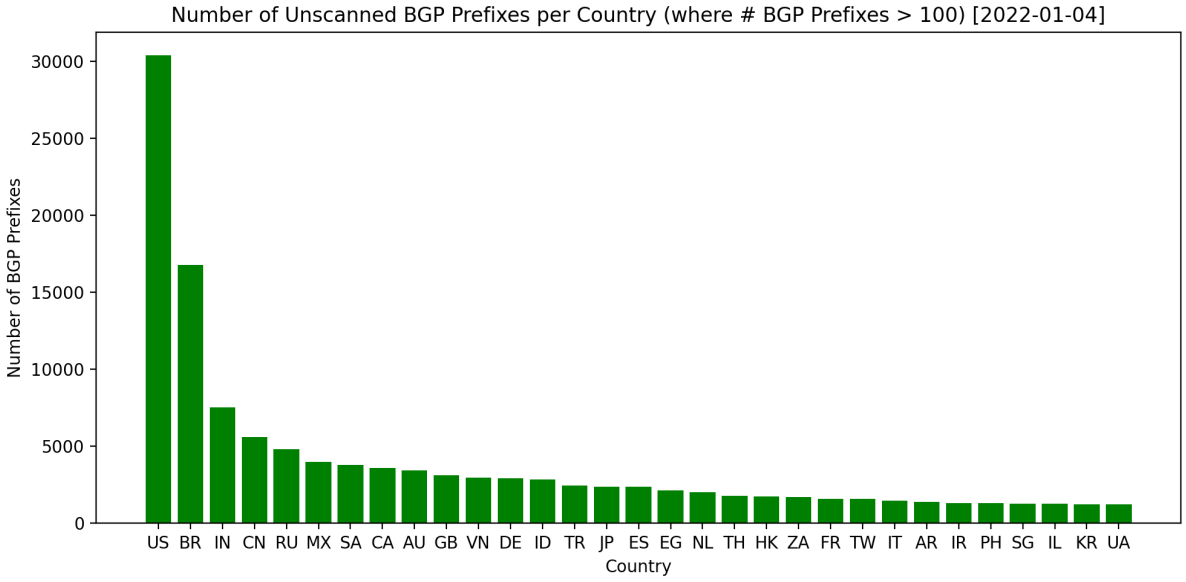


Figure 6.11: The number of non-scanned BGP prefixes per country, when the GeoIP2Location data returned multiple countries for a BGP prefix, it is counted towards each country.

The same approach was used for non-scanned BGP prefixes. Both figures show the number of BGP prefixes per country. To limit the number of countries shown in the figures, we limited the number of countries to countries that have more than 3000 for scanned prefixes and 1000 for non-scanned prefixes. Both for scanned and non-scanned BGP prefixes show the United States and Brazil as the top two countries.
An interesting observation is the number of prefixes of China. Coming in at the 12th place in scanned prefixes, but is at the 4th position in the non-scanned prefixes.

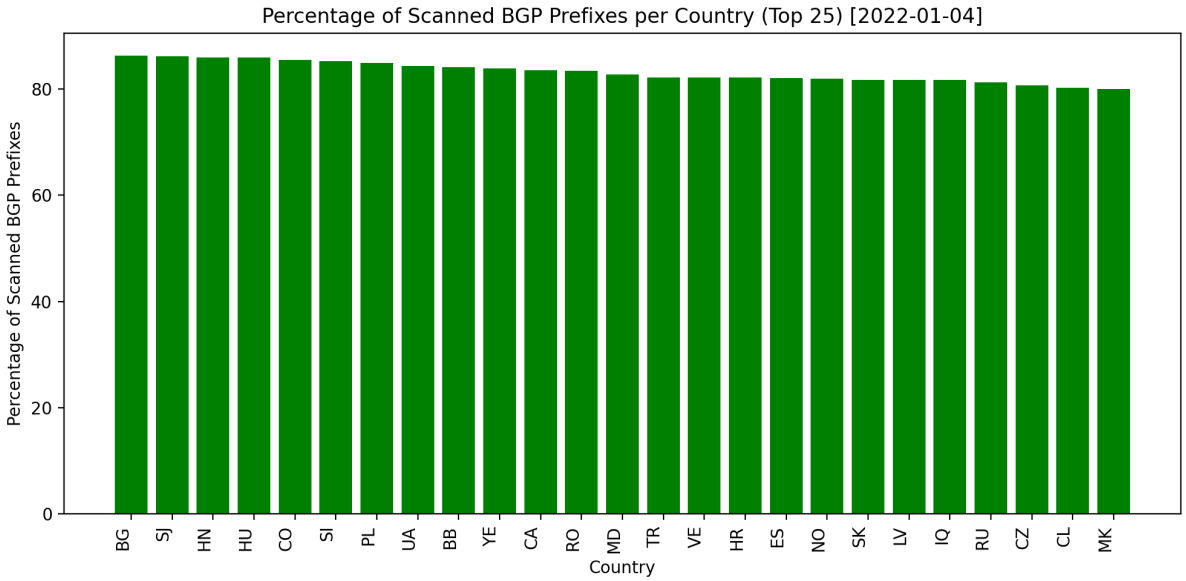


Figure 6.12: The percentage of BGP prefixes scanned per country - Top 25

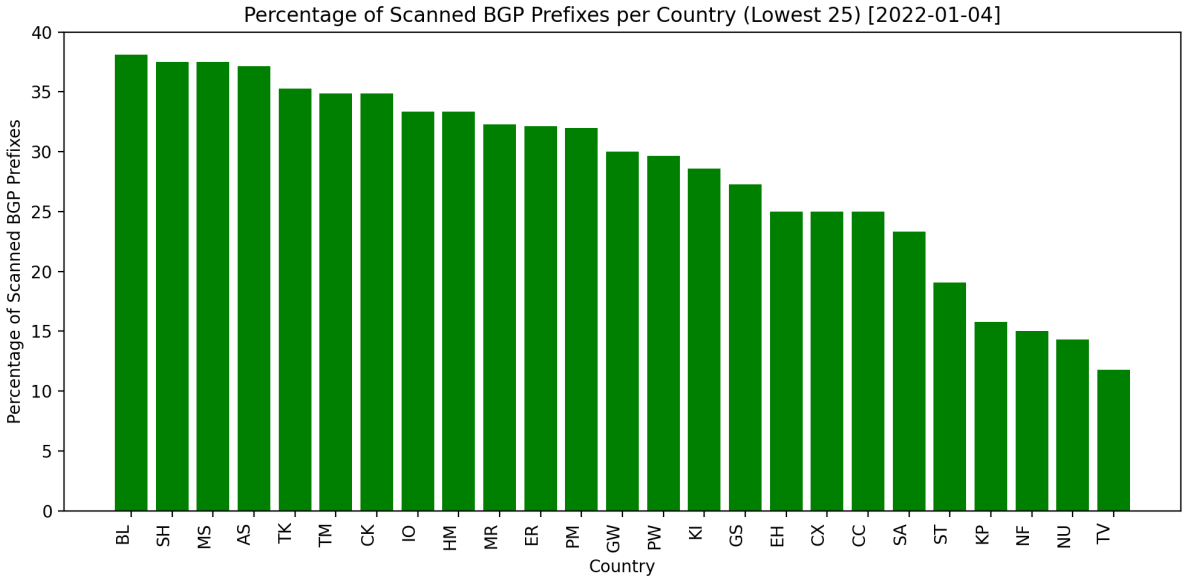


Figure 6.13: The percentage of BGP prefixes scanned per country - Lowest 25

Figure 6.12 and 6.13 shows the percentage of scanned BGP prefixes that are assigned to each country. Bulgaria is at the top with 86.23% of its prefixes scanned, while Tuvalu

has the lowest percentage of scanned prefixes.

See Figure 10.4 and 10.4 in the Appendix for full country names.

The full data is available in Table 10.4 in the Appendix. This shows the number of BGP prefixes per country, as well as the total number of assigned IP addresses and the number of IP addresses scanned.

An interesting observation is the great difference between North Korea and South Korea. Where North Korea has 19 prefixes and 29.049 IP addresses, South Korea has 5.646 prefixes and 113.913.642 IP addresses. Also only 0.10% (28 IP addresses) of North Korea is present in the scan, while South Korea has 10.13% of its IP addresses scanned (11.536.430)

## 6.7 Prefix Length

To get a better understanding whether prefix length matters, we analysed the percentage of scanned BGP prefix per prefix length.

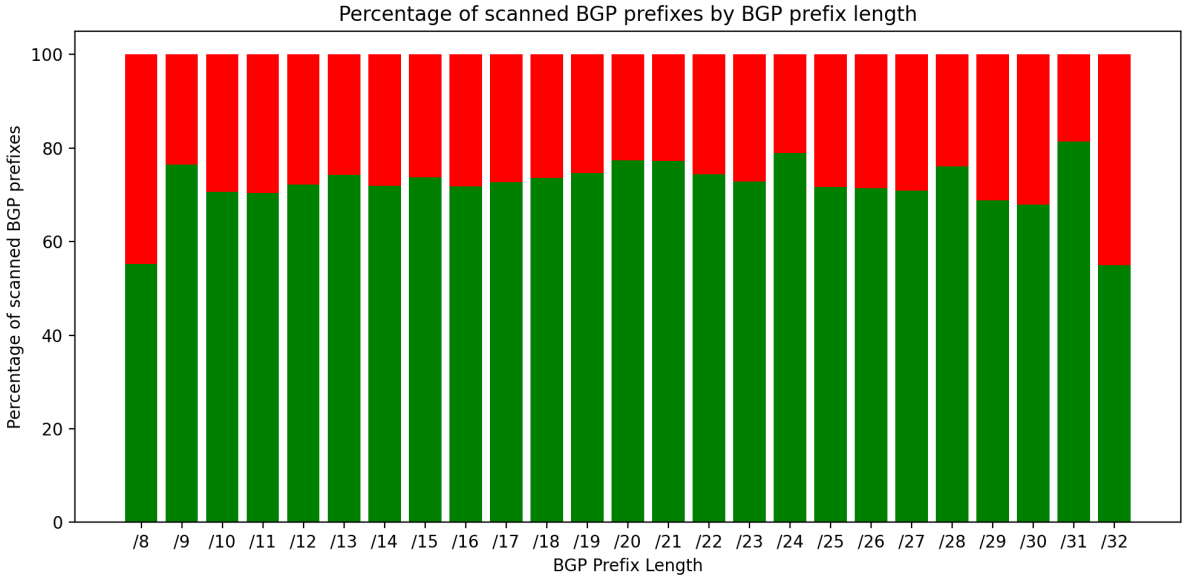


Figure 6.14: The percentage of scanned BGP prefixes per country, when the GeoIP2Location data returned multiple countries for a BGP prefix, it is counted towards each country.

As can be seen in Figure 6.14, the percentage of prefix length is quite stable, except for the smallest (/32) and largest (/8) prefix lengths. A prefix length of /32 only contains a single IP address and is therefore harder to scan. A more interesting observation is that multiple assigned /8 prefixes are not scanned at all. As can be seen in Table 6.1, a total of 11 /8 prefixes are not scanned. All except one are assigned to the Department of Defence of the United Stats of America, and therefore have a Military usage type. Another observation is the assigned /8 prefix to Daimler AG, currently Mercedes-Benz Group. In the early stages of the Internet, large prefixes were assigned to companies,

but were often redistributed later when IPv4 addresses were becoming scarce, except for this prefix assigned to Daimler which is currently not in use. Another large prefix assigned to a company is 17.0.0.0/8, which is assigned to Apple Inc., but this prefix was found to be scanned.

| Prefix | AS Organisation | Usage Type |
|---|---|---|
| 7.0.0.0/8 | DoD (Department of Defence) Ukraine and USA | MIL |
| 11.0.0.0/8 | DoD (Department of Defence) Brazil, China, Germany and USA | MIL |
| 21.0.0.0/8 | DoD (Department of Defence) USA | MIL |
| 22.0.0.0/8 | DoD (Department of Defence) Belgium and USA | MIL |
| 26.0.0.0/8 | DoD (Department of Defence) USA | MIL |
| 28.0.0.0/8 | DoD (Department of Defence) USA | MIL |
| 29.0.0.0/8 | DoD (Department of Defence) USA | MIL |
| 30.0.0.0/8 | DoD (Department of Defence) USA | MIL |
| 33.0.0.0/8 | DoD (Department of Defence) USA | MIL |
| 53.0.0.0/8 | Daimler AG (Mercedes-Benz Group AG) | COM |
| 55.0.0.0/8 | US Army | MIL |

Table 6.1: Large assigned /8 prefixes, without any scan results.

# 7 CONCLUSIONS

**[Q1.] What portion of the internet is not scanned?**
We can conclude that Censys' scans have a coverage of around 5.2% of the total IPv4 internet address space, which means 94.8% is not scanned. When we look at the number of ASs we can see that 96% of the ASs with at least 1 IP address is scanned, so 4% of the ASs are not scanned. When we look at the number of BGP prefixes that are scanned, we can see that 69% of the assigned BGP prefixes has at least 1 scanned IP address and 31% of the BGP prefixes is not scanned.

**[Q2.] What is the usage type of the networks that are not included in the Internet-wide scans?**
From the data we can conclude that BGP prefixes with the internet service provider usage type is the most non-scanned usage type in absolute numbers, following by that of data centers. When looking at percentage, we can see that only 44% of BGP prefixes with the military usage type is scanned, which means a majority of 56% of the military BGP prefixes is not scanned.

**[Q3.] What geographic part of the world is scanned and non-scanned?**
We can see that the United States has the most prefixes and is in both scanned and non-scanned prefix at the top, followed by Brazil. China has a relatively high number of non-scanned prefixes compared to the scanned prefix. When looking at the percentage, BGP prefixes in Bulgaria are scanned the most with 86% of its prefixes scanned and 14% is non-scanned. Least scanned is Tuvalu with only 11.8% of its BGP prefixes scanned.

# 8 DISCUSSION

## 8.1 Limitations

The dataset provided by Censys only contained scan results for the 3200 ports that they scan. Services that are active on other ports that are not scanned, are not discovered by Censys and therefore not counted as 'scanned IP address'.

Scans by Censys are executed over a period of time. A complete scan by Censys is roughly executed over the period of a week, making is more like an observation over a period of time instead of instant snapshot. Hosts might appear or disappear during this period.

Since Censys executes these scans every week on a large number of ports, it is likely that target hosts have taken measures to prevent this possibly unwanted traffic. Target hosts may have set up firewalls or intrusion detection systems may have flagged Censys' scanning IP addresses and blocked them.

As far as Censys discloses on its website at the time of writing, the scans are executed from servers in the United States only. As earlier mentioned in the Related Works section, the scanning location makes a difference on whether a host can be reached.

## 8.2 Data sources

This research builds upon the assumption that the data provided by IP2Location for classifying countries and usage type is valid. Inconsistencies are not filtered by any means from this dataset. IP2Location claims to have an accuracy of 99.5% for country-level detection. While this is accuracy is quite high, the accuracy is calculated by IP2Location itself.

IP2Location does not publicly disclose the sources they use for aggregating the data provided in the dataset. However, the site mentions that it "performs internal quality assurance by validating data quality against known IP addresses sourced from the public on a regular basis".

# 9 FUTURE WORK

Since this research uses only the data provided by Censys, the definition of coverage is limited to the scanning capabilities of Censys. Other Internet-wide scanning projects, such as those of Rapid7 [12], can possibly uncover a different set of reachable internet address ranges, due to different scanning methods, reporting, or location.

A possible future research could compare the data from Censys to Rapid7 to discover the differences between the results. This might be hard since the data that Rapid7 provides is not as structured as the data from Censys. While Censys does a complete scan at one point in time for all 3200 ports. Rapid7 irregularly scans on a more limited set of ports in different intervals than Censys.

Another possible approach is to start a scan in exactly the same way as Censys and compare the results. Scanning from a fresh IP range could reveal IP addresses and services that do not respond to Censys' scans, but do respond to the fresh IP range. When executed at the same time, from the same location, this will essentially result in a list of hosts that block the scans of Censys.

The country classification also includes small countries in terms of population. It would be interesting to see whether the data can be compared to other properties of the countries, like population, GDP, or Development Index. There might be a correlation between Development Index and the percentage of BGP prefixes and IP addresses that are scanned. (More development, more services, more reachable services, higher scan rates?)

# REFERENCES


[1] N. Provos and P. Honeyman, "ScanSSH - Scanning the Internet for SSH Servers," pp. 25–30, 12 2001. [Online]. Available: https://www.usenix.org/legacy/events/lisa01/tech/full_papers/provos/provos_html/

[2] "Nmap." [Online]. Available: https://nmap.org/

[3] "masscan." [Online]. Available: https://www.kali.org/tools/masscan/

[4] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide Scanning and Its Security Applications," in *22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX Association, 8 2013, pp. 605–620. [Online]. Available: https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/durumeric

[5] G. F. Lyon, *Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning*. Sunnyvale, CA, USA: Insecure, 2009.

[6] S. Bano, P. Richter, M. Javed, S. Sundaresan, Z. Durumeric, S. J. Murdoch, R. Mortier, and V. Paxson, "Scanning the Internet for Liveness," *SIGCOMM Comput. Commun. Rev.*, vol. 48, no. 2, pp. 2–9, 5 2018. [Online]. Available: https://doi.org/10.1145/3213232.3213234

[7] D. Adrian, Z. Durumeric, G. Singh, and J. A. Halderman, "Zippier ZMap: Internet-Wide Scanning at 10 Gbps," in *8th USENIX Workshop on Offensive Technologies (WOOT 14)*. San Diego, CA: USENIX Association, 8 2014. [Online]. Available: https://www.usenix.org/conference/woot14/workshop-program/presentation/adrian

[8] L. Izhikevich, R. Teixeira, and Z. Durumeric, "LZR: Identifying Unexpected Internet Services," in *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 8 2021, pp. 3111–3128. [Online]. Available: https://www.usenix.org/conference/usenixsecurity21/presentation/izhikevich

[9] L. Izhikevich, R. Teixeira, and Z. Durumeric, "Predicting IPv4 Services across All Ports," in *Proceedings of the ACM SIGCOMM 2022 Conference*, ser. SIGCOMM '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 503–515. [Online]. Available: https://doi.org/10.1145/3544216.3544249

[10] G. Wan, L. Izhikevich, D. Adrian, K. Yoshioka, R. Holz, C. Rossow, and Z. Durumeric, "On the Origin of Scanning: The Impact of Location on Internet-Wide Scans," in *Proceedings of the ACM Internet Measurement Conference*, ser.

IMC '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 662–679. [Online]. Available: https://doi.org/10.1145/3419394.3424214

[11] "Censys." [Online]. Available: https://censys.io/

[12] "Rapid7." [Online]. Available: https://www.rapid7.com/

[13] "Shodan.io." [Online]. Available: https://www.shodan.io/

[14] A. Murdock, F. Li, P. Bramsen, Z. Durumeric, and V. Paxson, "Target Generation for Internet-Wide IPv6 Scanning," in *Proceedings of the 2017 Internet Measurement Conference*, ser. IMC '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 242–253. [Online]. Available: https://doi.org/10.1145/3131365.3131405

[15] Censys, "Censys - Critical Vulnerability in OpenSSL!" 10 2022. [Online]. Available: https://censys.io/critical-vulnerability-in-openssl/

[16] M. Karl, M. Musch, G. Ma, M. Johns, and S. Lekies, "No Keys to the Kingdom Required: A Comprehensive Investigation of Missing Authentication Vulnerabilities in the Wild," in *Proceedings of the 22nd ACM Internet Measurement Conference*, ser. IMC '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 619–632. [Online]. Available: https://doi.org/10.1145/3517745.3561446

[17] A. Mirian, Z. Ma, D. Adrian, M. Tischer, T. Chuenchujit, T. Yardley, R. Berthier, J. Mason, Z. Durumeric, J. A. Halderman, and M. Bailey, "An Internet-wide view of ICS devices," in *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, 2016, pp. 96–103.

[18] L. Markowsky and G. Markowsky, "Scanning for vulnerable devices in the Internet of Things," in *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 1, 2015, pp. 463–467.

[19] J. O'Hare, R. Macfarlane, and O. Lo, "Identifying Vulnerabilities Using Internet-Wide Scanning Data," in *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, 2019, pp. 1–10.

[20] R. Holz, J. Hiller, J. Amann, A. Razaghpanah, T. Jost, N. Vallina-Rodriguez, and O. Hohlfeld, "Tracking the Deployment of TLS 1.3 on the Web: A Story of Experimentation and Centralization," *SIGCOMM Comput. Commun. Rev.*, vol. 50, no. 3, pp. 3–15, 7 2020. [Online]. Available: https://doi.org/10.1145/3411740.3411742

[21] O. Gasser, R. Holz, and G. Carle, "A deeper understanding of SSH: Results from Internet-wide scans," in *2014 IEEE Network Operations and Management Symposium (NOMS)*, 2014, pp. 1–9.

[22] Censys, "Censys - Internet Connectivity in Puerto Rico Post-Hurricane Fiona," 9 2022. [Online]. Available: https://censys.io/internet-connectivity-in-puerto-rico-post-hurricane-fiona/

[23] A. Feal, P. Vallina, J. Gamba, S. Pastrana, A. Nappa, O. Hohlfeld, N. Vallina-Rodriguez, and J. Tapiador, “Blocklist Babel: On the Transparency and Dynamics of Open Source Blocklisting,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1334–1349, 2021.

[24] Internet Assigned Numbers Authority, “IANA Homepage.” [Online]. Available: https://www.iana.org/

[25] University of Oregon RouteViews Project, “RouteViews.” [Online]. Available: http://archive.routeviews.org/

[26] “AS Report for UTWENTE-AS.” [Online]. Available: https://bgp.potaroo.net/cgi-bin/as-report?as=AS1133&view=2.0

[27] M. Ziv, L. Izhikevich, K. Ruth, K. Izhikevich, and Z. Durumeric, “ASdb: A System for Classifying Owners of Autonomous Systems,” in *Proceedings of the 21st ACM Internet Measurement Conference*, ser. IMC ’21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 703–719. [Online]. Available: https://doi.org/10.1145/3487552.3487853

[28] M. Luckie, A. Marder, M. Fletcher, B. Huffaker, and K. Claffy, “Learning to Extract and Use ASNs in Hostnames,” in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 386–392. [Online]. Available: https://doi.org/10.1145/3419394.3423639

[29] “Rapid7 Open Data.” [Online]. Available: https://opendata.rapid7.com/about/

[30] Censys, “Censys - Opt Out of Data Collection,” 8 2022. [Online]. Available: https://support.censys.io/hc/en-us/articles/360043177092-Opt-Out-of-Data-Collection

[31] S. Ramanathan, A. Hossain, J. Mirkovic, M. Yu, and S. Afroz, “Quantifying the Impact of Blocklisting in the Age of Address Reuse,” in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 360–369. [Online]. Available: https://doi.org/10.1145/3419394.3423657

[32] L. V. Guo, G. Akiwate, L. Kirill, V. G. M., and S. Stefan, “Clairvoyance: Inferring Blocklist Use on the Internet,” in *Passive and Active Measurement*, H. Oliver, A. Lutu, and L. Dave, Eds. Cham: Springer International Publishing, 2021, pp. 57–75.

[33] IANA, “IANA IPv4 Special-Purpose Address Registry,” 2 2021. [Online]. Available: http://www.iana.org/assignments/iana-ipv4-special-registry/iana-ipv4-special-registry.xhtml

[34] P. Richter and A. Berger, “Scanning the Scanners: Sensing the Internet from a Massively Distributed Network Telescope,” in *Proceedings of the Internet Measurement Conference*, ser. IMC ’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 144–157. [Online]. Available: https://doi.org/10.1145/3355369.3355595

[35] R. Elias, E. Glatz, D. Xenofonta, and D. Alberto, “How Dangerous Is Internet Scanning?” in *Traffic Monitoring and Analysis*, S. Moritz, P. Barlet-Ros, and B. Olivier, Eds. Cham: Springer International Publishing, 2015, pp. 158–172.

[36] H. Heo and S. Shin, “Who is Knocking on the Telnet Port: A Large-Scale Empirical Study of Network Scanning,” in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ser. ASIACCS ’18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 625–636. [Online]. Available: https://doi.org/10.1145/3196494.3196537

[37] H. N. Viet, Q. N. Van, L. L. T. Trang, and S. Nathan, “Using Deep Learning Model for Network Scanning Detection,” in *Proceedings of the 4th International Conference on Frontiers of Educational Technologies*, ser. ICFET ’18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 117–121. [Online]. Available: https://doi.org/10.1145/3233347.3233379

[38] “SURFcert’s Guidelines for Scans (in Dutch).” [Online]. Available: https://wiki.surfnet.nl/display/SURFcert/Richtlijnen+voor+scans

[39] M. Bailey, D. Dittrich, E. Kenneally, and D. Maughan, “The Menlo Report,” *IEEE Security & Privacy Magazine*, vol. 10, no. 2, pp. 71–75, 3 2012.

# 10 APPENDIX

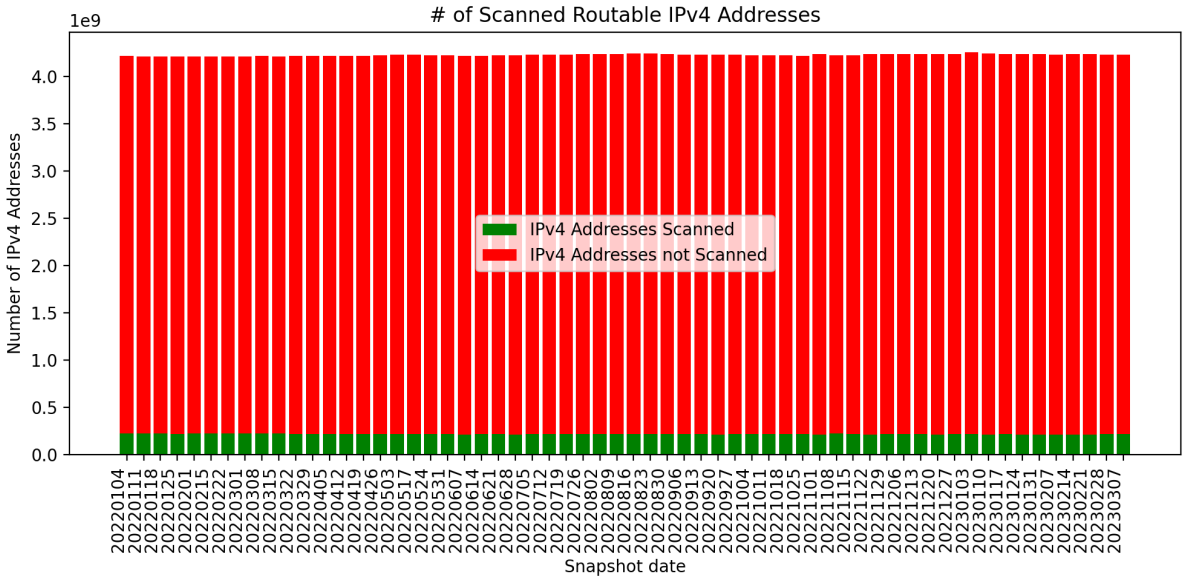


Figure 10.1: Number of total IP addresses and scanned IP address per Censys scan date

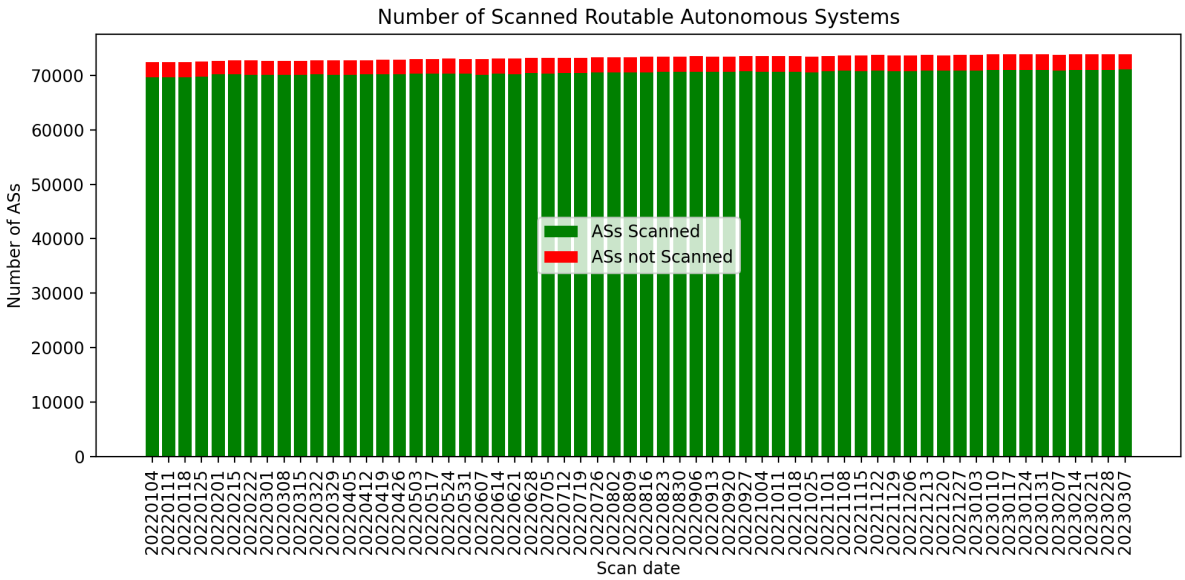


Figure 10.2: Number of total Autonomous Systems and scanned Autonomous Systems per Censys snapshot

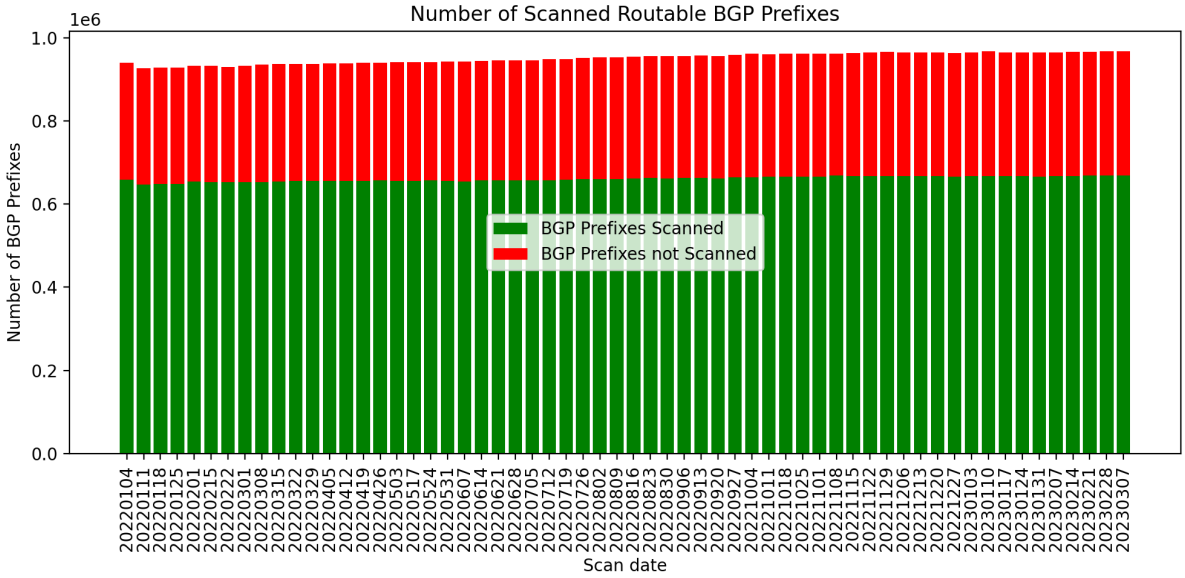


Figure 10.3: Number of total BGP Prefixes and scanned BGP Prefixes per Censys snapshot

| Range | Count |
|---|---|
| 0% - 5% | 36985 |
| 5% - 10% | 14529 |
| 10% - 15% | 6755 |
| 15% - 20% | 3697 |
| 20% - 25% | 2169 |
| 25% - 30% | 1426 |
| 30% - 35% | 977 |
| 35% - 40% | 625 |
| 40% - 45% | 460 |
| 45% - 50% | 372 |
| 50% - 55% | 311 |
| 55% - 60% | 200 |
| 60% - 65% | 169 |
| 65% - 70% | 129 |
| 70% - 75% | 99 |
| 75% - 80% | 111 |
| 80% - 85% | 68 |
| 85% - 90% | 124 |
| 95% - 100% | 309 |

Table 10.1: The percentage rangees of IP addresses scanned in ASs

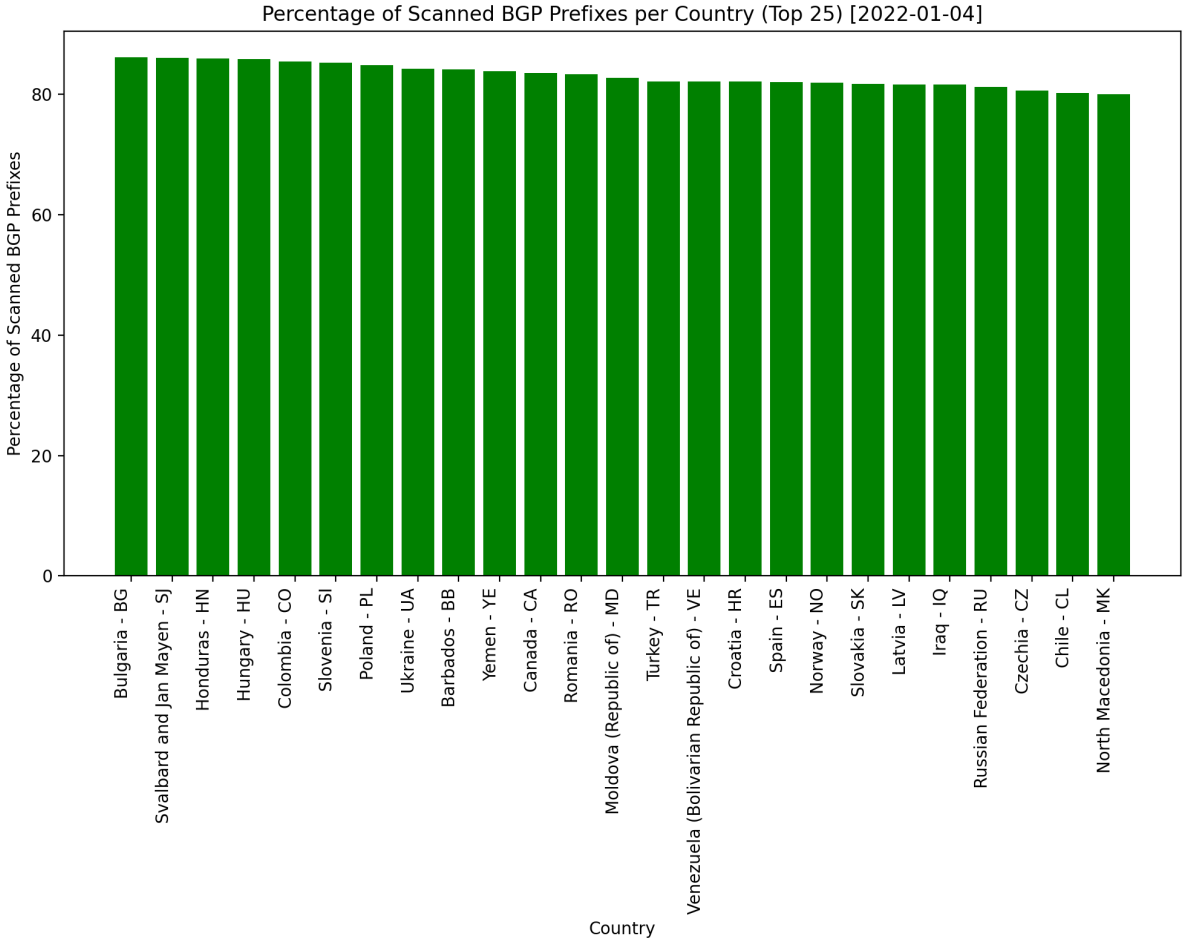


Figure 10.4: The percentage of BGP prefixes scanned per country - Top 25

| Range | Count |
|---|---|
| 0% - 5% | 276556 |
| 5% - 10% | 109492 |
| 10% - 15% | 57779 |
| 15% - 20% | 37435 |
| 20% - 25% | 26125 |
| 25% - 30% | 21749 |
| 30% - 35% | 16231 |
| 35% - 40% | 11767 |
| 40% - 45% | 8966 |
| 45% - 50% | 7521 |
| 50% - 55% | 7503 |
| 55% - 60% | 6560 |
| 60% - 65% | 5910 |
| 65% - 70% | 5449 |
| 70% - 75% | 5284 |
| 75% - 80% | 4983 |
| 80% - 85% | 4541 |
| 85% - 90% | 9564 |
| 95% - 100% | 29394 |

Table 10.2: The percentage ragens of IP addresses scanned in BGP Prefixes

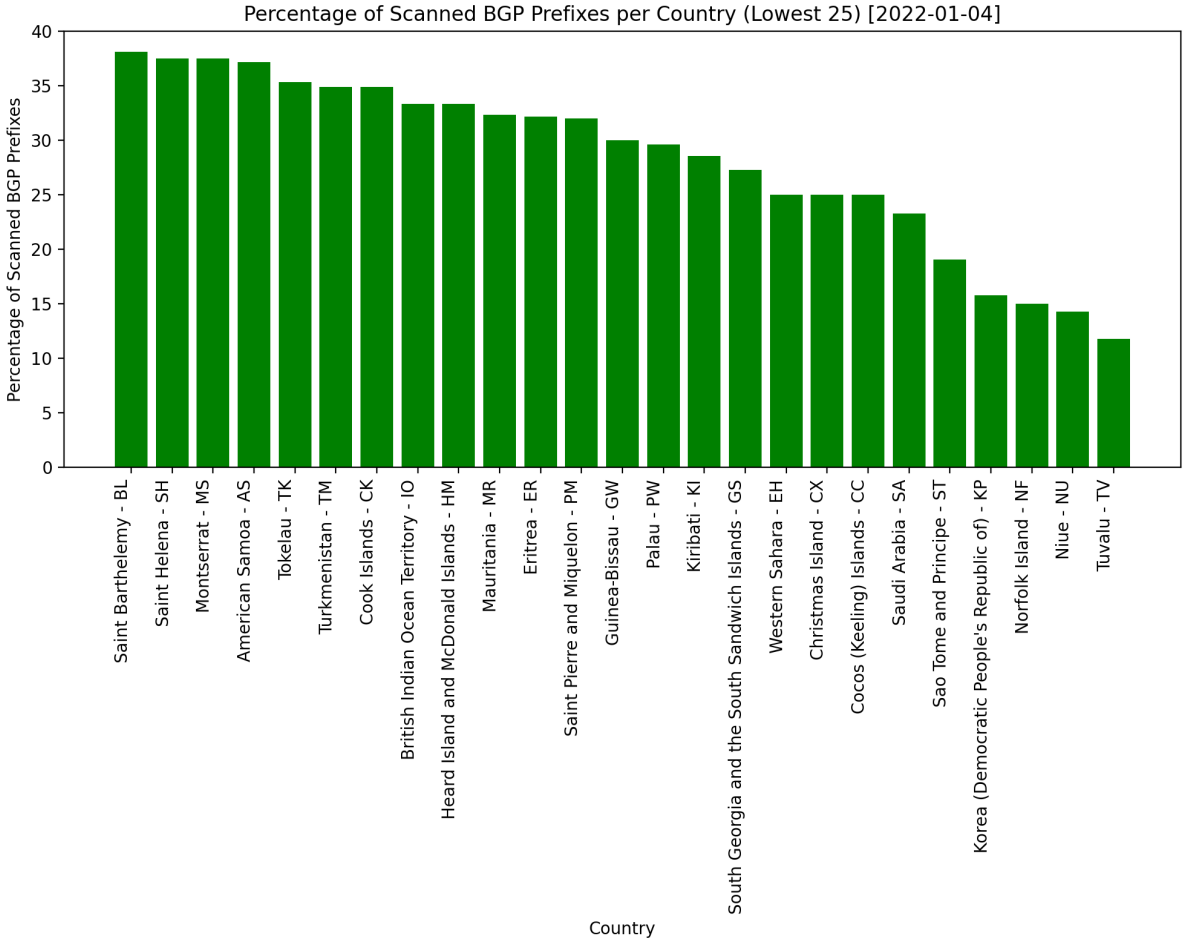


Figure 10.5: The percentage of BGP prefixes scanned per country - Lowest 25

| Usage Type | Abbreviation | Non-scanned | Scanned |
|---|---|---|---|
| Internet Service Provider (Fixed) | ISP | 54787 | 180125 |
| Internet Service Provider (Mobile & Fixed) | ISP/MOB | 32090 | 71519 |
| Data Center/Web Hosting/Transit | DHC | 24817 | 95299 |
| Commercial | COM | 23829 | 79092 |
| Internet Service Provider (Mobile) | MOB | 10798 | 16084 |
| Educational | EDU | 5521 | 19120 |
| Government | GOV | 2599 | 8947 |
| Organization | ORG | 2489 | 11119 |
| Content Delivery Network | CDN | 2445 | 7719 |
| Military | MIL | 1184 | 927 |
| Library | LIB | 239 | 1287 |
| Search Engine Spider | SES | 214 | 704 |
| Unknown | - | 29 | 56 |
| Reserved | RSV | 2 | 5 |

Table 10.3: The count of usage types of scanned and non-scanned BGP Prefixes

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Afghanistan - AF | 362 | 245 | 67.68 % | 246725 | 14077 | 5.71 % |
| Aland Islands - AX | 30 | 17 | 56.67 % | 48621 | 1291 | 2.66 % |
| Albania - AL | 467 | 330 | 70.66 % | 382319 | 48597 | 12.71 % |
| Algeria - DZ | 442 | 265 | 59.95 % | 4767706 | 235725 | 4.94 % |
| American Samoa - AS | 35 | 13 | 37.14 % | 7874 | 147 | 1.87 % |
| Andorra - AD | 105 | 77 | 73.33 % | 62201 | 22096 | 35.52 % |
| Angola - AO | 275 | 190 | 69.09 % | 1254675 | 17115 | 1.36 % |
| Anguilla - AI | 52 | 31 | 59.62 % | 10444 | 2717 | 26.01 % |
| Antarctica - AQ | 15 | 6 | 40.00 % | 578 | 256 | 44.29 % |
| Antigua and Barbuda - AG | 145 | 79 | 54.48 % | 48394 | 5842 | 12.07 % |
| Argentina - AR | 6914 | 5514 | 79.75 % | 19401835 | 4005371 | 20.64 % |
| Armenia - AM | 340 | 261 | 76.76 % | 640495 | 65764 | 10.27 % |
| Aruba - AW | 103 | 65 | 63.11 % | 126718 | 4441 | 3.50 % |
| Australia - AU | 16080 | 12635 | 78.58 % | 53027574 | 4279117 | 8.07 % |
| Austria - AT | 2681 | 2143 | 79.93 % | 11197872 | 662087 | 5.91 % |
| Azerbaijan - AZ | 454 | 330 | 72.69 % | 776327 | 42293 | 5.45 % |
| Bahamas - BS | 150 | 106 | 70.67 % | 160634 | 20977 | 13.06 % |
| Bahrain - BH | 483 | 307 | 63.56 % | 889316 | 57332 | 6.45 % |
| Bangladesh - BD | 3468 | 2567 | 74.02 % | 1823227 | 221143 | 12.13 % |
| Barbados - BB | 340 | 286 | 84.12 % | 146673 | 46448 | 31.67 % |
| Belarus - BY | 731 | 517 | 70.73 % | 1885503 | 122368 | 6.49 % |
| Belgium - BE | 2330 | 1798 | 77.17 % | 15435254 | 672635 | 4.36 % |
| Belize - BZ | 209 | 149 | 71.29 % | 138601 | 47421 | 34.21 % |
| Benin - BJ | 116 | 76 | 65.52 % | 155244 | 7872 | 5.07 % |
| Bermuda - BM | 107 | 72 | 67.29 % | 123727 | 7208 | 5.83 % |
| Bhutan - BT | 81 | 45 | 55.56 % | 41445 | 4074 | 9.83 % |
| Bolivia (Plurinational State of) - BO | 609 | 369 | 60.59 % | 1188244 | 87680 | 7.38 % |
| Bonaire - BQ | 39 | 23 | 58.97 % | 28778 | 1842 | 6.40 % |
| Bosnia and Herzegovina - BA | 586 | 453 | 77.30 % | 811515 | 113054 | 13.93 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Botswana - BW | 172 | 116 | 67.44 % | 171366 | 15177 | 8.86 % |
| Bouvet Island - BV | 9 | 7 | 77.78 % | 1356 | 569 | 41.96 % |
| Brazil - BR | 44112 | 27331 | 61.96 % | 89514650 | 7892261 | 8.82 % |
| British Indian Ocean Territory - IO | 15 | 5 | 33.33 % | 3648 | 89 | 2.44 % |
| Brunei Darussalam - BN | 152 | 91 | 59.87 % | 222524 | 11579 | 5.20 % |
| Bulgaria - BG | 4575 | 3945 | 86.23 % | 4280473 | 417736 | 9.76 % |
| Burkina Faso - BF | 172 | 130 | 75.58 % | 336113 | 8033 | 2.39 % |
| Burundi - BI | 83 | 34 | 40.96 % | 40373 | 1525 | 3.78 % |
| Cabo Verde - CV | 42 | 18 | 42.86 % | 32108 | 17166 | 53.46 % |
| Cambodia - KH | 987 | 526 | 53.29 % | 452636 | 72979 | 16.12 % |
| Cameroon - CM | 232 | 124 | 53.45 % | 670010 | 9696 | 1.45 % |
| Canada - CA | 21862 | 18262 | 83.53 % | 72625859 | 9618485 | 13.24 % |
| Cayman Islands - KY | 83 | 55 | 66.27 % | 65944 | 25210 | 38.23 % |
| Central African Republic - CF | 52 | 28 | 53.85 % | 9005 | 204 | 2.27 % |
| Chad - TD | 85 | 46 | 54.12 % | 31756 | 1002 | 3.16 % |
| Chile - CL | 2855 | 2292 | 80.28 % | 10440768 | 429191 | 4.11 % |
| China - CN | 14763 | 9162 | 62.06 % | 352313251 | 14581127 | 4.14 % |
| Christmas Island - CX | 4 | 1 | 25.00 % | 8 | 2 | 25.00 % |
| Cocos (Keeling) Islands - CC | 4 | 1 | 25.00 % | 8 | 2 | 25.00 % |
| Colombia - CO | 6181 | 5286 | 85.52 % | 17578884 | 949285 | 5.40 % |
| Comoros - KM | 26 | 11 | 42.31 % | 11614 | 410 | 3.53 % |
| Congo - CG | 97 | 67 | 69.07 % | 151046 | 1705 | 1.13 % |
| Congo (Democratic Republic of the) - CD | 242 | 170 | 70.25 % | 195395 | 4974 | 2.55 % |
| Cook Islands - CK | 43 | 15 | 34.88 % | 8861 | 276 | 3.11 % |
| Costa Rica - CR | 2341 | 1653 | 70.61 % | 2298385 | 132099 | 5.75 % |
| Cote d’Ivoire - CI | 545 | 323 | 59.27 % | 1687346 | 82587 | 4.89 % |
| Croatia - HR | 806 | 662 | 82.13 % | 2434898 | 92975 | 3.82 % |
| Cuba - CU | 82 | 38 | 46.34 % | 261009 | 6174 | 2.37 % |
| CuraÃ§ao - CW | 103 | 68 | 66.02 % | 177734 | 20808 | 11.71 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Cyprus - CY | 874 | 691 | 79.06 % | 966327 | 70527 | 7.30 % |
| Czechia - CZ | 2146 | 1731 | 80.66 % | 8725784 | 532611 | 6.10 % |
| Denmark - DK | 1898 | 1503 | 79.19 % | 12996351 | 391980 | 3.02 % |
| Djibouti - DJ | 60 | 30 | 50.00 % | 69257 | 732 | 1.06 % |
| Dominica - DM | 83 | 44 | 53.01 % | 37737 | 3685 | 9.76 % |
| Dominican Republic - DO | 619 | 481 | 77.71 % | 1635955 | 441550 | 26.99 % |
| Ecuador - EC | 2045 | 1520 | 74.33 % | 2785321 | 192083 | 6.90 % |
| Egypt - EG | 5367 | 3229 | 60.16 % | 24184989 | 339489 | 1.40 % |
| El Salvador - SV | 796 | 598 | 75.13 % | 697204 | 23332 | 3.35 % |
| Equatorial Guinea - GQ | 59 | 35 | 59.32 % | 23188 | 2125 | 9.16 % |
| Eritrea - ER | 28 | 9 | 32.14 % | 7289 | 204 | 2.80 % |
| Estonia - EE | 604 | 476 | 78.81 % | 1404806 | 92596 | 6.59 % |
| Eswatini - SZ | 74 | 52 | 70.27 % | 59443 | 7524 | 12.66 % |
| Ethiopia - ET | 183 | 91 | 49.73 % | 369630 | 13253 | 3.59 % |
| Falkland Islands (Malvinas) - FK | 25 | 10 | 40.00 % | 6355 | 167 | 2.63 % |
| Faroe Islands - FO | 41 | 20 | 48.78 % | 45693 | 9507 | 20.81 % |
| Fiji - FJ | 107 | 74 | 69.16 % | 150152 | 10526 | 7.01 % |
| Finland - FI | 1944 | 1514 | 77.88 % | 15274741 | 535063 | 3.50 % |
| France - FR | 7712 | 6126 | 79.43 % | 82986640 | 8291884 | 9.99 % |
| French Guiana - GF | 37 | 20 | 54.05 % | 108869 | 20423 | 18.76 % |
| French Polynesia - PF | 189 | 140 | 74.07 % | 73790 | 44809 | 60.73 % |
| French Southern Territories - TF | 5 | 3 | 60.00 % | 8 | 2 | 25.00 % |
| Gabon - GA | 94 | 62 | 65.96 % | 482040 | 2577 | 0.53 % |
| Gambia - GM | 38 | 22 | 57.89 % | 269670 | 3436 | 1.27 % |
| Georgia - GE | 448 | 325 | 72.54 % | 1262355 | 102040 | 8.08 % |
| Germany - DE | 14269 | 11333 | 79.42 % | 135636163 | 16252024 | 11.98 % |
| Ghana - GH | 424 | 311 | 73.35 % | 2258328 | 15615 | 0.69 % |
| Gibraltar - GI | 85 | 54 | 63.53 % | 81829 | 4230 | 5.17 % |
| Greece - GR | 1536 | 1166 | 75.91 % | 5748979 | 1260358 | 21.92 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Greenland - GL | 59 | 34 | 57.63 % | 35150 | 4158 | 11.83 % |
| Grenada - GD | 93 | 61 | 65.59 % | 33261 | 1985 | 5.97 % |
| Guadeloupe - GP | 78 | 52 | 66.67 % | 203528 | 50838 | 24.98 % |
| Guam - GU | 94 | 65 | 69.15 % | 233428 | 25964 | 11.12 % |
| Guatemala - GT | 592 | 452 | 76.35 % | 664080 | 32704 | 4.92 % |
| Guernsey - GG | 83 | 65 | 78.31 % | 67982 | 15822 | 23.27 % |
| Guinea - GN | 66 | 36 | 54.55 % | 37084 | 2416 | 6.51 % |
| Guinea-Bissau - GW | 30 | 9 | 30.00 % | 6900 | 238 | 3.45 % |
| Guyana - GY | 86 | 57 | 66.28 % | 72654 | 10734 | 14.77 % |
| Haiti - HT | 121 | 72 | 59.50 % | 161316 | 10753 | 6.67 % |
| Heard Island and McDonald Islands - HM | 3 | 1 | 33.33 % | 4 | 2 | 50.00 % |
| Holy See - VA | 22 | 11 | 50.00 % | 13943 | 244 | 1.75 % |
| Honduras - HN | 678 | 583 | 85.99 % | 484441 | 37952 | 7.83 % |
| Hong Kong - HK | 6234 | 4474 | 71.77 % | 20273903 | 4131438 | 20.38 % |
| Hungary - HU | 1645 | 1413 | 85.90 % | 6002620 | 338814 | 5.64 % |
| Iceland - IS | 296 | 217 | 73.31 % | 911409 | 63492 | 6.97 % |
| India - IN | 24613 | 17068 | 69.35 % | 46092156 | 2812399 | 6.10 % |
| Indonesia - ID | 10996 | 8131 | 73.95 % | 20013270 | 981262 | 4.90 % |
| Iran (Islamic Republic of) - IR | 5090 | 3759 | 73.85 % | 12380751 | 867024 | 7.00 % |
| Iraq - IQ | 1223 | 999 | 81.68 % | 815460 | 42775 | 5.25 % |
| Ireland - IE | 1832 | 1314 | 71.72 % | 14111111 | 2017032 | 14.29 % |
| Isle of Man - IM | 137 | 87 | 63.50 % | 127127 | 33401 | 26.27 % |
| Israel - IL | 5357 | 4072 | 76.01 % | 8310962 | 930614 | 11.20 % |
| Italy - IT | 6793 | 5318 | 78.29 % | 56598815 | 8485260 | 14.99 % |
| Jamaica - JM | 401 | 312 | 77.81 % | 276083 | 59820 | 21.67 % |
| Japan - JP | 9811 | 7437 | 75.80 % | 196718270 | 5430311 | 2.76 % |
| Jersey - JE | 79 | 51 | 64.56 % | 90314 | 21922 | 24.27 % |
| Jordan - JO | 975 | 665 | 68.21 % | 716960 | 117915 | 16.45 % |
| Kazakhstan - KZ | 3150 | 2410 | 76.51 % | 3326699 | 327488 | 9.84 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Kenya - KE | 1514 | 1057 | 69.82 % | 6241052 | 85023 | 1.36 % |
| Kiribati - KI | 35 | 10 | 28.57 % | 6330 | 251 | 3.97 % |
| Korea (Democratic People's Republic of) - KP | 19 | 3 | 15.79 % | 29049 | 28 | 0.10 % |
| Korea (Republic of) - KR | 5646 | 4406 | 78.04 % | 113913642 | 11536430 | 10.13 % |
| Kuwait - KW | 1061 | 480 | 45.24 % | 1959393 | 66526 | 3.40 % |
| Kyrgyzstan - KG | 265 | 163 | 61.51 % | 281719 | 16441 | 5.84 % |
| Lao People's Democratic Republic - LA | 210 | 105 | 50.00 % | 92694 | 9754 | 10.52 % |
| Latvia - LV | 683 | 558 | 81.70 % | 1923074 | 206121 | 10.72 % |
| Lebanon - LB | 883 | 613 | 69.42 % | 622791 | 37947 | 6.09 % |
| Lesotho - LS | 70 | 28 | 40.00 % | 128068 | 2414 | 1.88 % |
| Liberia - LR | 81 | 42 | 51.85 % | 108926 | 1128 | 1.04 % |
| Libya - LY | 182 | 143 | 78.57 % | 436468 | 7894 | 1.81 % |
| Liechtenstein - LI | 116 | 87 | 75.00 % | 124345 | 13624 | 10.96 % |
| Lithuania - LT | 1698 | 1324 | 77.97 % | 2963562 | 226247 | 7.63 % |
| Luxembourg - LU | 619 | 485 | 78.35 % | 1627904 | 93993 | 5.77 % |
| Macao - MO | 116 | 55 | 47.41 % | 388338 | 52091 | 13.41 % |
| Madagascar - MG | 226 | 99 | 43.81 % | 583070 | 4200 | 0.72 % |
| Malawi - MW | 181 | 114 | 62.98 % | 554888 | 5071 | 0.91 % |
| Malaysia - MY | 2192 | 1400 | 63.87 % | 6898118 | 576879 | 8.36 % |
| Maldives - MV | 159 | 94 | 59.12 % | 93973 | 5189 | 5.52 % |
| Mali - ML | 76 | 46 | 60.53 % | 92372 | 21207 | 22.96 % |
| Malta - MT | 498 | 334 | 67.07 % | 678969 | 42336 | 6.24 % |
| Marshall Islands - MH | 33 | 17 | 51.52 % | 7080 | 1977 | 27.92 % |
| Martinique - MQ | 65 | 42 | 64.62 % | 171056 | 44264 | 25.88 % |
| Mauritania - MR | 65 | 21 | 32.31 % | 49437 | 1821 | 3.68 % |
| Mauritius - MU | 521 | 247 | 47.41 % | 6781389 | 28132 | 0.41 % |
| Mayotte - YT | 26 | 12 | 46.15 % | 29556 | 6652 | 22.51 % |
| Mexico - MX | 14098 | 10124 | 71.81 % | 29524057 | 1545538 | 5.23 % |
| Micronesia (Federated States of) - FM | 41 | 16 | 39.02 % | 14564 | 1137 | 7.81 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Moldova (Republic of) - MD | 586 | 485 | 82.76 % | 1265328 | 182543 | 14.43 % |
| Monaco - MC | 73 | 50 | 68.49 % | 62222 | 10318 | 16.58 % |
| Mongolia - MN | 285 | 177 | 62.11 % | 197126 | 16885 | 8.57 % |
| Montenegro - ME | 122 | 75 | 61.48 % | 193178 | 8110 | 4.20 % |
| Montserrat - MS | 32 | 12 | 37.50 % | 3590 | 109 | 3.04 % |
| Morocco - MA | 2047 | 1345 | 65.71 % | 12273427 | 961514 | 7.83 % |
| Mozambique - MZ | 525 | 267 | 50.86 % | 470948 | 14607 | 3.10 % |
| Myanmar - MM | 461 | 252 | 54.66 % | 219846 | 10630 | 4.84 % |
| Namibia - NA | 134 | 87 | 64.93 % | 439007 | 38549 | 8.78 % |
| Nauru - NR | 26 | 13 | 50.00 % | 14479 | 126 | 0.87 % |
| Nepal - NP | 742 | 461 | 62.13 % | 592567 | 26579 | 4.49 % |
| Netherlands - NL | 9207 | 7195 | 78.15 % | 52790851 | 3081895 | 5.84 % |
| New Caledonia - NC | 57 | 40 | 70.18 % | 165966 | 24752 | 14.91 % |
| New Zealand - NZ | 2445 | 1813 | 74.15 % | 7284463 | 955648 | 13.12 % |
| Nicaragua - NI | 287 | 217 | 75.61 % | 443631 | 30155 | 6.80 % |
| Niger - NE | 88 | 46 | 52.27 % | 48568 | 2981 | 6.14 % |
| Nigeria - NG | 1352 | 928 | 68.64 % | 3178013 | 101601 | 3.20 % |
| Niue - NU | 21 | 3 | 14.29 % | 2140 | 162 | 7.57 % |
| Norfolk Island - NF | 20 | 3 | 15.00 % | 2136 | 31 | 1.45 % |
| North Macedonia - MK | 535 | 428 | 80.00 % | 719025 | 46008 | 6.40 % |
| Northern Mariana Islands - MP | 33 | 19 | 57.58 % | 17788 | 1387 | 7.80 % |
| Norway - NO | 1991 | 1632 | 81.97 % | 15788806 | 271747 | 1.72 % |
| Oman - OM | 559 | 239 | 42.75 % | 1023662 | 106527 | 10.41 % |
| Pakistan - PK | 2827 | 1943 | 68.73 % | 6176127 | 574311 | 9.30 % |
| Palau - PW | 27 | 8 | 29.63 % | 11452 | 409 | 3.57 % |
| Palestine - PS | 813 | 569 | 69.99 % | 849794 | 150551 | 17.72 % |
| Panama - PA | 952 | 728 | 76.47 % | 1769085 | 104723 | 5.92 % |
| Papua New Guinea - PG | 150 | 99 | 66.00 % | 79810 | 4642 | 5.82 % |
| Paraguay - PY | 599 | 433 | 72.29 % | 1254811 | 124571 | 9.93 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Peru - PE | 1967 | 1193 | 60.65 % | 3464552 | 127164 | 3.67 % |
| Philippines - PH | 3432 | 2109 | 61.45 % | 6196751 | 273855 | 4.42 % |
| Pitcairn - PN | 5 | 2 | 40.00 % | 12 | 3 | 25.00 % |
| Poland - PL | 6307 | 5355 | 84.91 % | 21342608 | 2128311 | 9.97 % |
| Portugal - PT | 856 | 662 | 77.34 % | 6810827 | 573081 | 8.41 % |
| Puerto Rico - PR | 953 | 741 | 77.75 % | 1250615 | 81787 | 6.54 % |
| Qatar - QA | 338 | 221 | 65.38 % | 916298 | 42936 | 4.69 % |
| Reunion - RE | 105 | 83 | 79.05 % | 509484 | 77293 | 15.17 % |
| Romania - RO | 4087 | 3408 | 83.39 % | 7742349 | 612742 | 7.91 % |
| Russian Federation - RU | 25706 | 20880 | 81.23 % | 45555796 | 5236282 | 11.49 % |
| Rwanda - RW | 175 | 106 | 60.57 % | 313176 | 7372 | 2.35 % |
| Saint Barthelemy - BL | 21 | 8 | 38.10 % | 2692 | 137 | 5.09 % |
| Saint Helena - SH | 8 | 3 | 37.50 % | 16 | 2 | 12.50 % |
| Saint Kitts and Nevis - KN | 99 | 60 | 60.61 % | 40311 | 6436 | 15.97 % |
| Saint Lucia - LC | 107 | 69 | 64.49 % | 41444 | 1535 | 3.70 % |
| Saint Martin (French Part) - MF | 59 | 33 | 55.93 % | 36862 | 5632 | 15.28 % |
| Saint Pierre and Miquelon - PM | 25 | 8 | 32.00 % | 6022 | 281 | 4.67 % |
| Saint Vincent and the Grenadines - VC | 70 | 42 | 60.00 % | 26769 | 1518 | 5.67 % |
| Samoa - WS | 55 | 30 | 54.55 % | 26263 | 891 | 3.39 % |
| San Marino - SM | 52 | 26 | 50.00 % | 40459 | 2175 | 5.38 % |
| Sao Tome and Principe - ST | 21 | 4 | 19.05 % | 14964 | 303 | 2.02 % |
| Saudi Arabia - SA | 4970 | 1158 | 23.30 % | 10791219 | 241030 | 2.23 % |
| Senegal - SN | 203 | 115 | 56.65 % | 409057 | 134486 | 32.88 % |
| Serbia - RS | 1130 | 896 | 79.29 % | 2407382 | 132239 | 5.49 % |
| Seychelles - SC | 224 | 172 | 76.79 % | 292964 | 69646 | 23.77 % |
| Sierra Leone - SL | 287 | 144 | 50.17 % | 116659 | 1441 | 1.24 % |
| Singapore - SG | 4153 | 2864 | 68.96 % | 17510009 | 2456224 | 14.03 % |
| Sint Maarten (Dutch Part) - SX | 33 | 19 | 57.58 % | 35855 | 2349 | 6.55 % |
| Slovakia - SK | 820 | 670 | 81.71 % | 2823749 | 153624 | 5.44 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Slovenia - SI | 684 | 583 | 85.23 % | 2666002 | 284731 | 10.68 % |
| Solomon Islands - SB | 32 | 14 | 43.75 % | 20052 | 701 | 3.50 % |
| Somalia - SO | 100 | 62 | 62.00 % | 55211 | 1680 | 3.04 % |
| South Africa - ZA | 4899 | 3171 | 64.73 % | 28241766 | 1136007 | 4.02 % |
| South Georgia / South Sandwich Islands - GS | 11 | 3 | 27.27 % | 292 | 3 | 1.03 % |
| South Sudan - SS | 73 | 43 | 58.90 % | 23282 | 2052 | 8.81 % |
| Spain - ES | 13228 | 10860 | 82.10 % | 35312263 | 2144553 | 6.07 % |
| Sri Lanka - LK | 335 | 134 | 40.00 % | 583961 | 59736 | 10.23 % |
| Sudan - SD | 298 | 124 | 41.61 % | 1892233 | 17717 | 0.94 % |
| Suriname - SR | 131 | 64 | 48.85 % | 83303 | 37162 | 44.61 % |
| Svalbard and Jan Mayen - SJ | 216 | 186 | 86.11 % | 576 | 96 | 16.67 % |
| Sweden - SE | 3897 | 3045 | 78.14 % | 30132997 | 1328430 | 4.41 % |
| Switzerland - CH | 3641 | 2860 | 78.55 % | 24379006 | 962367 | 3.95 % |
| Syrian Arab Republic - SY | 291 | 130 | 44.67 % | 1297295 | 31925 | 2.46 % |
| Taiwan (Province of China) - TW | 3694 | 2123 | 57.47 % | 36461582 | 2462900 | 6.75 % |
| Tajikistan - TJ | 169 | 94 | 55.62 % | 88279 | 5147 | 5.83 % |
| Tanzania - TZ | 545 | 380 | 69.72 % | 1090123 | 21673 | 1.99 % |
| Thailand - TH | 6753 | 4978 | 73.72 % | 9534345 | 1421939 | 14.91 % |
| Timor-Leste - TL | 46 | 18 | 39.13 % | 19182 | 2021 | 10.54 % |
| Togo - TG | 80 | 51 | 63.75 % | 355953 | 7940 | 2.23 % |
| Tokelau - TK | 17 | 6 | 35.29 % | 3726 | 20 | 0.54 % |
| Tonga - TO | 28 | 11 | 39.29 % | 16214 | 401 | 2.47 % |
| Trinidad and Tobago - TT | 347 | 242 | 69.74 % | 536091 | 30788 | 5.74 % |
| Tunisia - TN | 670 | 335 | 50.00 % | 7876933 | 529161 | 6.72 % |
| Turkey - TR | 13794 | 11335 | 82.17 % | 16859620 | 1452789 | 8.62 % |
| Turkmenistan - TM | 43 | 15 | 34.88 % | 20948 | 1765 | 8.43 % |
| Turks and Caicos Islands - TC | 37 | 21 | 56.76 % | 14390 | 5830 | 40.51 % |
| Tuvalu - TV | 17 | 2 | 11.76 % | 12888 | 48 | 0.37 % |
| Uganda - UG | 512 | 325 | 63.48 % | 1446645 | 22596 | 1.56 % |

| | BGP prefixes | | | IP addresses | | |
|---|---|---|---|---|---|---|
| Country | Total | Scanned | Percentage | Total | Scanned | Percentage |
| Ukraine - UA | 7806 | 6581 | 84.31 % | 11108114 | 1100920 | 9.91 % |
| United Arab Emirates - AE | 1093 | 779 | 71.27 % | 4108897 | 264968 | 6.45 % |
| United Kingdom - GB | 13934 | 10828 | 77.71 % | 116822995 | 18747259 | 16.05 % |
| United States Minor Outlying Islands - UM | 20 | 9 | 45.00 % | 526 | 222 | 42.21 % |
| United States of America - US | 131023 | 100655 | 76.82 % | 1559413025 | 51131284 | 3.28 % |
| Uruguay - UY | 683 | 456 | 66.76 % | 2458488 | 557785 | 22.69 % |
| Uzbekistan - UZ | 403 | 252 | 62.53 % | 344961 | 19572 | 5.67 % |
| Vanuatu - VU | 48 | 29 | 60.42 % | 26241 | 806 | 3.07 % |
| Venezuela (Bolivarian Republic of) - VE | 1541 | 1266 | 82.15 % | 6932070 | 410259 | 5.92 % |
| Viet Nam - VN | 8622 | 5656 | 65.60 % | 16226361 | 1176891 | 7.25 % |
| Virgin Islands (British) - VG | 186 | 132 | 70.97 % | 55537 | 15283 | 27.52 % |
| Virgin Islands (U.S.) - VI | 156 | 118 | 75.64 % | 128940 | 2327 | 1.80 % |
| Wallis and Futuna - WF | 17 | 7 | 41.18 % | 3666 | 101 | 2.76 % |
| Western Sahara - EH | 4 | 1 | 25.00 % | 4 | 2 | 50.00 % |
| Yemen - YE | 155 | 130 | 83.87 % | 240266 | 38814 | 16.15 % |
| Zambia - ZM | 224 | 124 | 55.36 % | 1696635 | 16823 | 0.99 % |
| Zimbabwe - ZW | 227 | 147 | 64.76 % | 187040 | 15028 | 8.03 % |

Table 10.4: The number and percentage of scanned BGP Prefixes and IP addresses per country