Classifying the Network Capacity of the Dutch IPv4 Address Space

Pim Beune
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
p.f.beune@student.utwente.nl

ABSTRACT

Hosts connected to the Internet have a wide diversity of network capacities/bandwidths. As these capacity variations introduce a large impact on Quality of Service, they influence the way people make use of the Internet. While usage-based measurements have been conducted to classify IP-addresses, bandwidth-based classifications are missing. This paper explores the bandwidth performance of the Dutch IPv4 address space by creating a mapping of the capacity distribution. To achieve this goal we have leveraged the speed test measurements of Measurement Lab, which is a platform where users can test their bandwidth performance. Considering the Coefficient of Variation, we have determined that the measurement data of Measurement Lab is sufficiently accurate to determine the bandwidth of a host. Furthermore, although we find that there is no general correlation between bandwidth and geographical location, there is a five-fold difference in network capacity among some Autonomous Systems.

Keywords

Geolocation, Network capacity, Autonomous System, Bandwidth

1. INTRODUCTION

The network capacities, i.e., the network throughputs of hosts linked to the Internet vary greatly. This is mainly due to the capacity difference between various physical transmission media. For example, while fiber optic cables allow for speeds in the range of 100 to 200 Gbps, twisted pair cables support up to 10 Gbps, coaxial cables are limited to a maximum speed of 10 Mbps and satellite Internet transmission has an average rate of 1 Mbps [1, 2]. Another reason could be the geographical location of the host as shown by Farrington et al. [3]. They found that in all of the British deep rural sampled areas, the highest broadband speed (17.4 Mbps) was below the average in urban areas. Besides, these differences could be related to Autonomous Systems (ASes). An AS is a set of IP routing prefixes under the control of a network operator that provides a common, clearly defined Internet routing policy [4]. Any machine or device that links to the Internet is connected through an AS. Since the different ASes are run by different organizations whose infrastructure could differ, the bandwidths of the hosts connected to them may vary. In addition, upstream ASes play a major role, as they could become a bottleneck. For example, the AS called *SURFnet* (`AS1103`) is the upstream provider of the AS of the University of Twente (`AS1133`). If *SURFnet* provides a maximum download speed of x Mbps, the download speed of the AS of the University of Twente will also be x Mbps at most. Finally, another cause of Internet bandwidth variations could be due to the fact that the host supports a lower bandwidth than the Internet service provider allows, which might be due to (residential) hardware limitations of the host itself.

These bandwidth variations can have a big impact on Quality of Service (QoS). Consequently, this influences the extent to which people use the Internet. A 2015 study found that those with connection speeds less than 3.5 Mbps were less likely to attempt "data-heavy" practices such as downloading, gaming and content creation, such as video. They also were less likely to participate in online social networking [3].

Though usage-based measurements have been performed to classify IP addresses (see Section 2), capacity classifications are yet to be conducted. Thus, the main goal of this paper is to explore the bandwidth diversity of the IPv4 address space. As an example we have only explored the Dutch IPv4 address space. However, as a possible future work other IP spaces can be explored as well (see Section 7). Our study is based on the following Research Questions (RQ) in order to accomplish this goal:

- **RQ1** How consistent are the speed measurement datasets in determining the network capacity of a host?
- **RQ2** What is the correlation between the network capacity and the geographic location of the hosts?
- **RQ3** Are high-bandwidth hosts clustered in some autonomous systems?

In order to address these research questions we have parsed the network performance measurements of Measurement Lab (M-Lab) [5] and geolocation dataset of MaxMind (GeoLite2) [6] (Section 4). The outcome of this research can potentially be leveraged to improve the QoS, by improving Internet speeds of certain areas first.

In this research we present the following contributions: Considering the Coefficient of Variation, we conclude that the measurement data of M-Lab is accurate enough to determine the network capacity of a single host. Furthermore, we determine that there is no clear association between the geographic location and network capacity of hosts in the Netherlands, despite the fact that several cities

35th Twente Student Conference on IT Jul. 2nd, 2021, Enschede, The Netherlands.

have an above average bandwidth. Finally, we discover that the bandwidth difference across ASes in the Netherlands can be of around five orders of magnitude.

The structure of this paper is as follows. In Section 2, we will present other work related to classifying IP addresses. The datasets and methodology to answer the research questions will be discussed in Section 3 and 4 respectively. Following, the results and discussion will be presented in Section 5 and the limitations thereof in Section 6. Finally, we conclude our paper in Section 8.

2. RELATED WORK

There have been various projects and research regarding classifying Internet addresses. Firstly, Dainotti et al. [7] looked at the possibility of passive network traffic measurements (e.g., analyzing internet traffic datasets) to evaluate IPv4 address space use. Passive measurements add no network traffic overhead and may be applicable to IPv6 as well. However, this approach faces two challenges: the limited visibility of any single viewpoint of traffic and the presence of spoofed IP addresses in packets, which can significantly skew results by implying fake addresses are active. To solve the first challenge they considered all IP addresses in a single /24 address-block to be active if at least one IP address in this block shows activity. For the second challenge, they created and tested a method for removing spoofed traffic from datasets obtained on both darknets and live networks, and discovered that the filtered data could be used to perform measurements of IP address space usage.

In 2019, Du et al. [8] developed a method named *FENet*. This is a trained neural network that can classify the type of IP address usage (e.g., DNS, gaming, streaming, social, apps, email) in a given network traffic dataset. Their work demonstrates a higher level of accuracy and consistency in classification in comparison to public data sets. Our research differs from these works in the sense that rather than doing the classification on the usage type basis, we aim to perform a network capacity based classification.

Finally, numerous Internet measurement platforms have emerged over the past few years that have deployed thousands of probes in different places around the world. In 2015, Bajpai and Schönwälder [9] presented a taxonomy of these platforms with two main categories: topology and performance analysis. The performance measurement platforms were further categorised depending on their deployment use-case: landline and mobile network measurements. They explore the coverage, scope, timeline, deployed measurement tools, and overall research impact of the Internet measurement platforms in detail.

One of the platforms noted in this taxonomy is Measurement Lab [5] (see Section 3). The data collected on M-Lab will be of great use for this research as their measurements contain (among other features) TCP throughput and available bandwidth rates.

Another platform that provides network speed measurements is Speedtest by *Ookla* [10]. Their global index reports that the average download speed in the Netherlands is around 157 Mbps. Comparing this to the index of M-Lab [11], the average download speed is approximately 55 Mbps. As the raw Speedtest data is commercial, we do not have access to it and thus, this difference can not be analyzed. Nevertheless, one potential cause for this bandwidth difference could be that the measurement infrastructure of both organizations are different. Another possible cause could be that Speedtest is more well-known than M-Lab, as M-Lab is a more research-oriented platform, while Speedtest is more consumer-oriented. Thus, it is possible that Speedtest has significantly more measurement data than M-Lab.

3. DATASETS

This section elaborates on the datasets that have been used in this study.

Measurement Lab (M-Lab) is a free, collaborative server network that enables researchers to deploy active Internet measuring tools. The M-Lab platform is capable of performing active network measurement tests [5]. Moreover, M-Lab performs measurements between a client and their measurement servers to investigate the end-to-end performance throughout the complete path. These measurements result in data created by users who perform tests on a voluntary basis using either the M-Lab site or through clients from third parties. The M-Lab tool we have used in this research is the Network Diagnostic Tool (NDT). With this tool, TCP throughput is measured between a client operating on the user's host and an M-Lab server. Data from tests is delivered in both directions. Metadata including client-specific information, such as OS type and version, is also gathered. For our research, the most relevant data this tool collects is listed as follows:

`ClientIP` The client IP address that performed the measurement. This field is relevant for this research for it denotes the IP (and thus the location) of the network performance test.

`C2S.MeanThroughputMbps` The measured client to server speed speed that the server reports. This data is important as it stands for the upload speed of a client.

`S2C.MeanThroughputMbps` The measured server to client speed that the server reports. This information is relevant because it is the download speed of a client.

`StartTime` The time and date the measurement started in UTC format. This field will be necessary for our research as it allows us to filter out IPs that do not have measurements for a certain number of distinct days.

In order to discover geolocation information about a certain IP address we have utilised the MaxMind GeoLite2 IP database [6], which is updated weekly. This dataset provides details such as the country, region or state city, latitude, longitude and ZIP code of origin for almost the entire IPv4 address space. Moreover, the GeoLite2 database comes with a number of APIs so that the data can be integrated in other software projects.

To identify which IP address belongs to which AS, we have used the Border Gateway protocol (BGP) dataset from the *University of Oregon Route Views Project* [12], which is a collection of real-time information regarding the global routing system of numerous Autonomous Systems throughout the Internet.

4. METHODOLOGY

This section outlines the steps we have taken to answer to each of the research questions.

4.1 Consistency of speed measurements

To answer the first research question, we will be parsing the M-Lab Network Diagnostic Tool (NDT) dataset [13] (see Section 3) to extract appropriate features like upload

Table 1. Statistics of the top 10 IP addresses with the most measurements

	Upload				Download			
Number of measurements	Mean	Median	Standard deviation	CV	Mean	Median	Standard deviation	CV
144904	20.83	20.22	3.99	0.19	21.50	21.00	4.16	0.19
63983	4.90	2.79	4.05	0.83	8.73	7.36	4.15	0.48
40557	12.62	10.59	5.76	0.46	29.29	18.28	23.18	0.79
921	48.37	49.87	5.44	0.11	178.99	107.38	97.59	0.55
405	39.50	40.22	2.68	0.07	252.81	83.25	214.10	0.85
396	36.12	40.25	8.55	0.24	478.31	501.29	78.78	0.16
396	229.16	93.16	161.92	0.71	229.04	92.77	164.39	0.72
395	39.84	39.99	1.92	0.05	484.20	489.74	56.16	0.12
395	8.78	8.82	0.52	0.06	43.26	43.78	2.77	0.06
394	33.00	37.73	9.16	0.28	474.63	494.84	64.52	0.14

and download speeds. After that, we will investigate the statistical metrics of the measurement data such as the median, mean μ, standard deviation σ and Coefficient of Variation (CV). The coefficient of variation, also known as the relative standard deviation, is used to measure the dispersion of a frequency or probability distribution, and can be calculated using Equation 1. The CV is used in various fields of science, as the standard deviation of data always needs to be understood in context of the mean of the data. Furthermore, the CV allows the dispersion of one dataset to be compared with the variation of another dataset, regardless of whether these sets apply the same unit of measurement. For example, the CV allows the variance in apple weights to be compared to the dispersion in tree height in a certain area.

$$\mathrm{CV} = \frac{\sigma}{\mu} \tag{1}$$

Following, we can infer the consistency of the measurement data. Please note that it may be possible that some IP addresses will temporarily show up in the NDT dataset, so to reduce the number of outlying measurements we will only be using IP addresses that have appeared for at least 90% of the timeline of the processed dataset.

4.2 Geolocation-based throughput distribution

To answer the second research question, we will repeat the same methodology as used in the first research question. However, this time we do not apply the constraint that every IP has to have occurred for at least 90% of the timeline (see Section 5.2). This methodology results in a dataset of IP addresses and their corresponding bandwidths. After that, we will take the IP addresses inside of the this dataset to map those measurements to a physical location, using the MaxMind GeoLite2 [6] dataset. Finally, we will be able to determine whether there is a correlation between network capacity and geographic location of the hosts.

4.3 AS-level throughput diversity

To answer the third research question we will be using a Python module called `pyasn` that allows for very quick IP address to Autonomous System Number (ASN) lookups. Under the hood, this module uses data from the *Route Views Project* Border Gateway Protocol dataset [12] to perform these lookups (see Section 3). There is a number of other datasets and tools to find an ASN for an IP, but we opted to use `pyasn` for its convenience and the fact that the processing for the first and second research question was done in Python as well.

Firstly, we will process the measurements of M-Lab's NDT dataset [13] (see Section 3) so that the capacity measurement data and the corresponding IP address can be extracted. Secondly, we will map these measurements to an AS using the `pyasn` tool. From this we can determine whether high-bandwidth hosts may be clustered for some autonomous systems.

5. RESULTS AND DISCUSSION

This section presents the results and the discussion from the methodology (Section 4).

5.1 Speed measurement data consistency

Due to time constraints, not all data from the NDT dataset has been parsed. The processed data is collected from 3 February 2021 to 14 May 2021 (this is a range of 100 days). After parsing this dataset as discussed in the methodology (see Section 4.1), we can calculate the mean, median and standard deviation per IP of the measurements conducted in the Netherlands. Because of the possibility that some IP addresses temporarily show up in the dataset, only those IP addresses that have appeared for at least 90 out of 100 days of the dataset will be considered, to reduce the number of outlying datapoints in this experiment. From this emerged a dataset of 175 unique Dutch IP addresses. The results can be found in Table 1; it lists statistics of the top 10 IPs in the Netherlands that have the most upload and download speed measurements. Figure 1 shows a Cumulative Distribution Function (CDF) of the CVs of these 175 IP addresses in the Netherlands. A CDF is a graph that allows the distribution of a dataset to be read. For example, in Figure 1, it can be seen that around 80% (0.8 on the y-axis) of the upload speed measurements have a CV of 0.2 (x-axis) or lower.

It can be seen that the download speeds are on average more dispersed than the upload speeds, which could be attributed to *bandwidth asymmetry*; these are connections where download speeds are often orders of magnitude larger than upload speeds. Asymmetric bandwidths can be found in (both wired and wireless) modem and satellite networks [14]. Inspecting Figure 1 with for example, a threshold CV of 0.3, it turns out that 86.9% of the download speed measurements are below this threshold, for the upload speeds this is 88.6%. Although choosing this CV threshold depends on the use-case, it can be concluded that the speed measurements are consistent in determining the network capacity of an individual host.

5.2 Geolocation and throughput correlation

To answer the second research question, we repeated the

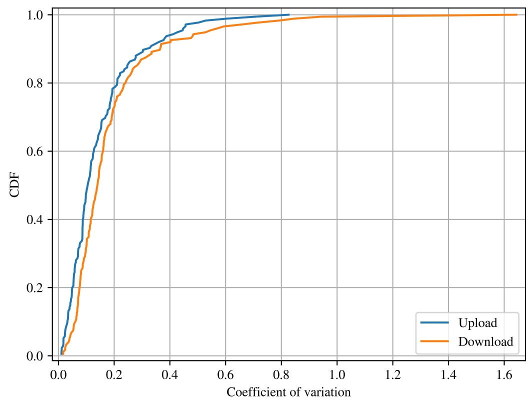

Figure 1. Coefficients of variation of the upload and download speed measurements by M-Lab

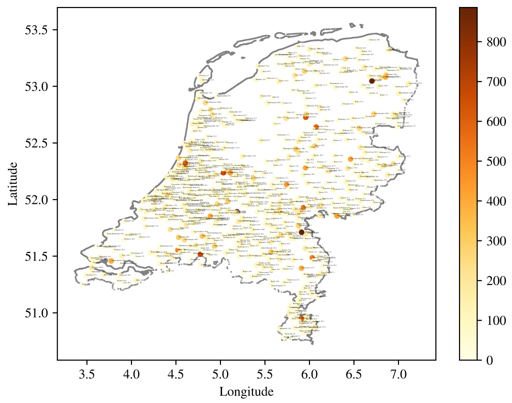

Figure 2. Download speed per city in the Netherlands (Mbps)

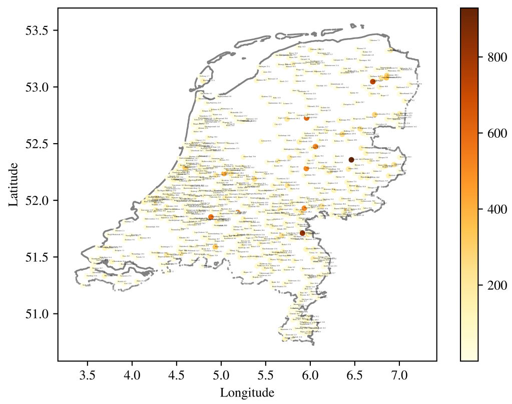

Figure 3. Upload speed per city in the Netherlands (Mbps)

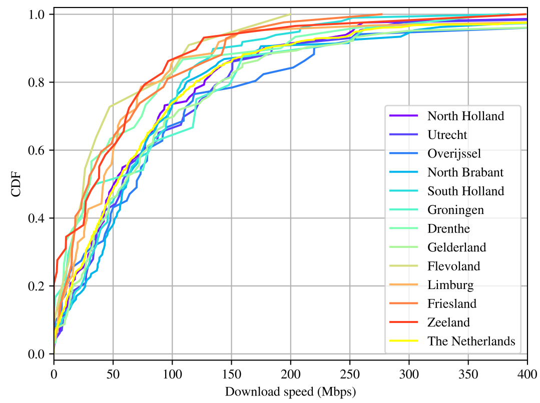

Figure 4. Download speed (Mbps) per city in the Netherlands, grouped by province

same methodology as discussed in RQ1, but without the 90% timeline constraint (see Section 4.1), as the data could be considered consistent regardless of this constraint (see Section 6). We end up with a dataset that contains around 760.000 speed measurements from 6858 unique Dutch IPv4 addresses. Aggregating this data per Dutch city results in a set of 680 unique cities.

Although the geographical plots in Figures 2 and 3 show that there are some cities with an on average high bandwidth, there is no clear correlation visible between the geographic location and network capacity of the hosts. Additionally, the geographical plots reveal that the measurements are spread out across the entire land area of the Netherlands. When looking at the CDFs in Figures 4 and 5 (x-axes have been limited to improve readability), it becomes clear that Drenthe and Friesland have an average low bandwidth in comparison with the rest of the Netherlands, which may be due to the fact that these provinces are considered to be more rural than other provinces. This could confirm the findings by Farrington et al. [3] for the Netherlands, as they found that rural areas in the United Kingdom have a below average bandwidth in comparison with urban areas. However, more work has to be done to confirm this (see Section 7). Inspecting provinces such as Overijssel and Gelderland, it can be seen that these areas have an on average higher bandwidth than that of the Netherlands.

Looking at the top 10 cities with the highest down- and upload speed (Tables 2 and 3), it becomes apparent that not all cities in the top 10 upload speeds are also in the top 10 download speeds. The cities to be included in both tables are Beesd, Anloo, Sint Agatha, Vreeland, Blokzijl and Huissen. In other words, the top 10 down- and upload speeds have an overlap of 60%. Furthermore, the bandwidth difference between the last and first city in the top 10 download speeds is about 65.6%, for the top 10 upload speeds this is roughly 125.2%. The higher dispersion in upload speeds could again be explained by *asymmetric bandwidths* (see Section 5.1). Another cause for the dispersion could be that ISPs may limit upload bandwidths to the point where the speeds are even lower than the hardware can handle.

5.3 High bandwidth AS clusters

To answer our third research question, the parsed data from RQ2 has been reused, which is a mapping of IP addresses and their respective bandwidths. Grouping the IPs in this mapping by AS using the BGP dataset [12], results in an aggregation of 124 unique Dutch ASes, with on average, around 54 distinct IP addresses per AS. It is important to note that there is a large variation between the number of IPs per AS; while there are 20 ASes that had just a single IP that conducted a measurement, there

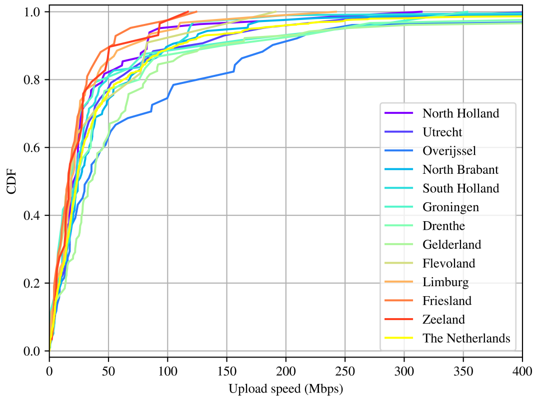

Figure 5. Upload speed (Mbps) per city in the Netherlands, grouped by province

Table 2. Top 10 cities in the Netherlands with the highest download speed

City	Province	Average download speed (Mbps)
Beesd	Gelderland	885.91
Anloo	Drenthe	881.92
Sint Agatha	North Brabant	878.59
Galder	North Brabant	668.79
Vreeland	Utrecht	636.87
Blokzijl	Overijssel	623.12
Bennebroek	North Holland	594.78
Oirsbeek	Limburg	586.79
Zwartsluis	Overijssel	553.80
Huissen	Gelderland	535.15

Table 3. Top 10 cities in the Netherlands with the highest upload speed

City	Province	Average upload speed (Mbps)
Nijverdal	Overijssel	928.76
Sint Agatha	North Brabant	816.66
Anloo	Drenthe	738.10
Blokzijl	Overijssel	616.01
Giessenburg	South Holland	568.32
Vaassen	Gelderland	513.36
Beesd	Gelderland	508.45
Huissen	Gelderland	507.13
Hattem	Gelderland	500.64
Vreeland	Utrecht	412.44

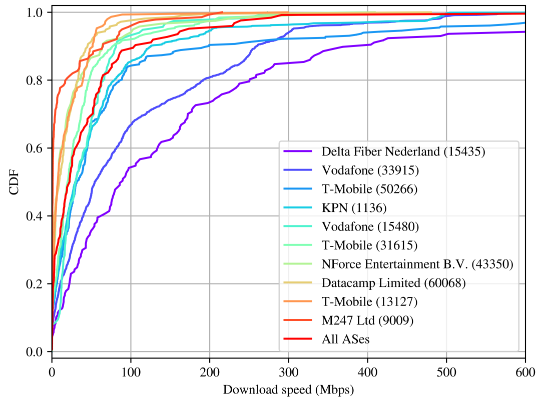

Figure 6. Top 10 ASes with the highest download speed in the Netherlands.

is one AS that had measurements from 1402 unique IPs.

In order to give an accurate representation of the bandwidth variation between ASes, we only included ASes with at least 100 unique IPs (and thus 100 different up and down measurements). Applying this constraint resulted in a dataset of 10 unique ASes. When looking at the CDFs, it becomes clear that there is a significant difference between the speed distribution of the top 10 ASes in the Netherlands. Figures 6 and 7 show that at some points, there is more than a 200 Mbps difference in up- and download speeds between the last and first AS in the top 10. For example, 20% of the IPs of *T-Mobile* (`AS31615`) have an upload speed of 50 Mbps or higher, while 20% of the IP addresses of *Delta Fiber Nederland* (`AS15435`) have an upload speed of 250 Mbps or higher. Inspecting the upstream ASes of the top AS in the dataset (*Delta Fiber Nederland*, `AS15435`), it appears that one of its three upstreams is *Stichting NBIP*, which is a foundation that provides supporting services to Internet providers. An example of such a service is DDoS attack protection, for which a huge bandwidth is needed.

The diversity in AS bandwidths could potentially be explained by the hierarchy of ASes; one AS can be the upstream provider of another AS. If the upstream provider provides limited bandwidth capabilities, due to hardware limitations for example, the downstream AS will have a limited bandwidth as well. When only looking at the ASes in the CDF, it appears that `AS33915` is the upstream provider for `AS15480`. One could claim that this is reflected in both the upload and download CDFs as the speeds of `AS15480` are lower than those of `AS33915`. However, it is important to note that in this research, we are exploring throughput variations on an IP-level. Therefore, we do not have any knowledge about the global throughput of ASes. Moreover, we do not have any insight in the provider-client (upstream-downstream) relations in ASes. Consequently, further research needs to be done in order to find the cause for these AS bandwidth variations (see Section 7).

6. LIMITATIONS

This section discusses the limitations of the results presented in Section 5.

To reduce the number of outliers of the experiment in the first research question, only the IPs that have performed a measurement for at least 90 unique days are included. This restriction brings a potential risk as the results may be biased towards a certain type of user; those who frequently monitor their bandwidth speed may have a high bandwidth in the first place. This may be possible because those who have a high link speed may want to keep it this way, and could complain to their ISP of their connection falls short. However, in Figure 8, where every datapoint denotes the average download speed of a single IP, numerous variations of this 90-day constraint are shown. As can be seen, apart from 10 days, most variations seem to follow the same distribution. Following, it could be assumed that the dataset is mostly consistent regardless of this constraint.

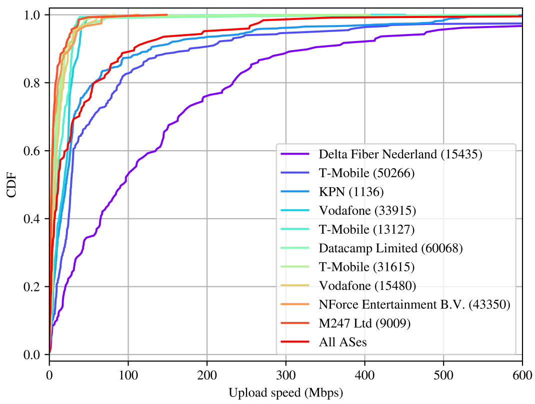

Figure 7. Top 10 ASes with the highest upload speed in the Netherlands.

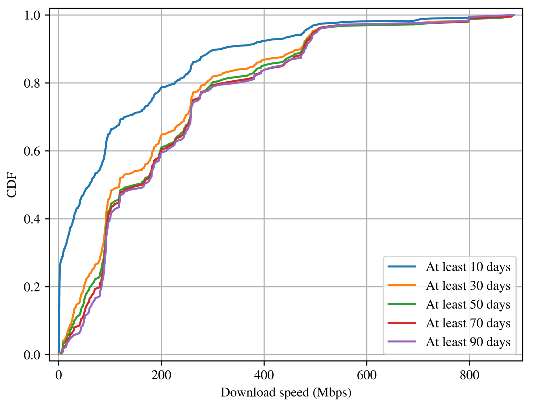

Figure 8. Variations of the 90-day constraint that was applied for the dataset in RQ1

As can be seen in the CDF in Figure 9, around 80% of the cities in the Netherlands have 200 download speed measurements or less. The same holds for uplink measurements. This number could be considered to be low, and thus to be insufficient for this research. Moreover, it is important to note that some Dutch IP addresses have conducted thousands of measurements, while others had just one. These aspects could both have an impact on the final results. Another parameter that may have an impact is the range of the data measurement period. As mentioned earlier, this range is 100 days due to time limitations of the research. As of writing the NDT dataset contains measurements dating back to July 2019, meaning that only approximately 15% of all NDT data has been processed. Furthermore, after aggregating the data for the second research question, we were left with a dataset of around 7000 unique Dutch IPv4 addresses, spread out over approximately 700 distinct cites. This means that the average number of IPv4 addresses per city is around 10. For this reason, the dataset can be considered to be limited and this should be taken into account when interpreting the findings of the research.

As mentioned earlier, to give an accurate representation of the bandwidth variation between ASes, only the ASes with at least 100 unique IPs (and thus 100 different up and down measurements) are included in the results. We did

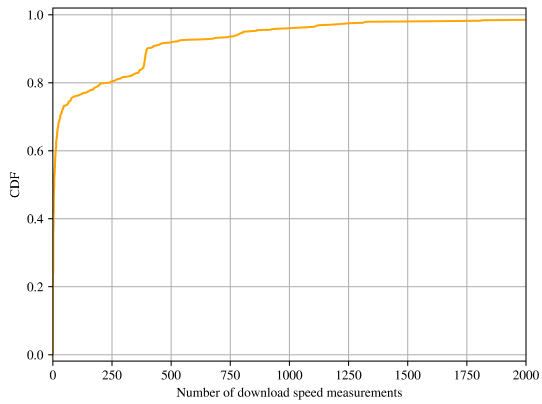

Figure 9. Number of download speed measurements per city in the Netherlands

Table 4. Provinces and their number of cities that have performed at least one M-Lab measurement

Province	Number of cities
North Brabant	106
Gelderland	103
South Holland	98
North Holland	82
Limburg	61
Overijssel	51
Utrecht	43
Friesland	42
Drenthe	30
Zeeland	29
Groningen	24
Flevoland	11

not apply a similar constraint for the data in RQ2 (e.g., at least x number of cities per province), as we were already working with only 12 provinces. Nevertheless, it can be seen in Table 4 that there are only two provinces that have more than 100 distinct cities with at least one M-Lab measurement. Furthermore, there are several provinces that have a considerably low number of cities. Therefore, it could be possible that the results for a number of provinces are unreliable as they only have a couple of cities with measurements. When interpreting the results from the second research question this fact should be taken into account.

Finally, the coverage of the IP data from M-Lab is unknown. Because M-Lab is a more research-oriented rather than consumer-oriented platform, it could be the case that more rural areas have performed significantly less bandwidth measurements on this platform than urban areas. This coverage can have an impact on the results of this research and should be investigated (see Section 7). This unknown coverage could result in another bias, where cities that have an above average bandwidth are more likely to perform bandwidth measurements. Nevertheless, inspecting the down- and upload speeds of the bottom 10 cities in Table 2 and 3, it turns out that their down- and uplink speeds range from 0.02–0.50 and 0.03–0.36 Mbps respectively. These bandwidths can be considered to be very low, meaning that low-bandwidth cities perform M-Lab measurements as well.

7. FUTURE WORK

To extend on this research, we propose the following ideas:

Firstly, instead of the IPv4 space, the IPv6 space could be classified. However, as of now the adoption of IPv6 is not as prevalent as that of IPv4 [15, 16]. Therefore, the representativeness of that measurement data could be unreliable.

Secondly, other IP address spaces can be classified. For example, the United States IP address space, as the US has many more datapoints than the Netherlands. For example, when processing M-Lab NDT data of a single day, 56.6% of the measurements originated from the United States while only 1.6% originated from the Netherlands. Consequently, the results of this future work could be more detailed than those in this study.

Additionally, the IPv4 and/or IPv6 coverage of the M-Lab measurement data could be examined to find out which subset of all IP addresses have performed a bandwidth measurement at M-Lab. To make this research computationally feasible, it could be assumed that all IP addresses under a specific address-block (e.g., /24) have similar bandwidths. Then, the IP addresses measurements can be grouped by such an address-block.

Another possible future work is investigating the cause of the bandwidth differences between cities/provinces and autonomous systems, as shown in RQ2 and RQ3 respectively. To do this, much more data has to be gathered than what has been accumulated for this research, as the current number of datapoints do not give a full view of the bandwidth of the cities and ASes. Furthermore, to find out this cause, it must be researched how different bandwidth distribution policies of the ASes affect their clients. To investigate the AS bandwidth relations other datasets can be used, like *CAIDA*'s *ASRank* project [17]. This project aims to create an autonomous system ranking on the basis of their influence in the global routing system. The three main metrics are the total number of ASes in the downstream paths (customer cone), prefixes and number of addresses in the AS.

Finally, using similar methodology as performed in this research, the study of Farrington et al. [3] could be repeated for the Dutch IP space. In their research, they systematically examined the bandwidth characteristics of urban, 'shallow' and 'deep rural' areas and investigated the characteristics of the (significant) minority of the British population who live across the British land area. They discovered that none of the deep rural respondents surveyed in a sample of about 1000 rural residents across the United Kingdom had access to fast broadband (>30 Mbps). They also found that the fastest Internet speed in any of the deep rural areas studied (17.4 Mbps) was slower than the average speed in urban areas [3].

8. CONCLUSIONS

Hosts with Internet access have a wide range of network capacities/bandwidths. In this paper we built a mapping of the Dutch IPv4 address space to explore its bandwidth variations. This mapping collects addresses that are in the same geolocation or AS and provides their respective throughput. To obtain this mapping we have used the throughput measurement data from Measurement Lab, the geolocation dataset from MaxMind and the BGP dataset from the *Route Views Project*. With this data we present the following findings:

After calculating statistical metrics of M-Lab data of IPv4 addresses in the Netherlands, we concluded that the speed measurement data from M-Lab is consistent and can thus be used to determine the network capacity of a single host.

Although we demonstrate that some cities have on average high bandwidth, we have discovered that in general, on a city level, no evident relationship between the geographic location and host network capacity could be discovered from the processed datasets. Furthermore, we have discovered that provinces that are considered to be rural have an on average lower bandwidth than other provinces. Nevertheless, whether this is a causal relation needs to be investigated further. In addition, it was discovered that the cities in the top 10 download and upload speeds have a reasonable overlap.

Investigating the AS speed CDFs, it is evident that the bandwidth difference between ASes can be of multiple orders of magnitude. From this, we conclude that there is a considerable variation in the speed distribution of the top ten ASes in the Netherlands. Nonetheless, to analyze why these differences occur, more research needs to be done.

References

[1] Fiber Optic Cable vs Twisted Pair Cable vs Coaxial Cable | FS Community, 2013. URL `https://community.fs.com/blog/the-difference-between-fiber-optic-cable-twisted-pair-and-cable.html`. Accessed: 2021-06-11.

[2] Everything You Need To Know About Coaxial Cable | RS Components. URL `https://uk.rs-online.com/web/generalDisplay.html?id=ideas-and-advice/coaxial-cable-guide`. Accessed: 2021-06-11.

[3] John Farrington, Lorna Philip, C. Cottrill, Pamela Abbott, Grant Blank, and William H. Dutton. Two-Speed Britain: Rural Internet Use. *SSRN Electronic Journal*, 2015. doi: 10.2139/ssrn.2645771.

[4] John Hawkinson and Tony Bates. Guidelines for Creation, Selection, and Registration of an Autonomous System (AS). BCP 6, RFC Editor, March 1996.

[5] Constantine Dovrolis, Krishna Gummadi, Aleksandar Kuzmanovic, and Sascha D. Meinrath. Measurement Lab: Overview and an Invitation to the Research Community. *ACM SIGCOMM Computer Communication Review*, 40(3):53–56, June 2010. doi: 10.1145/1823844.1823853.

[6] GeoLite2 Free Geolocation Data | MaxMind Developer Portal. URL `https://dev.maxmind.com/geoip/geolite2-free-geolocation-data`. Accessed: 2021-06-15.

[7] Alberto Dainotti, Karyn Benson, Alistair King, kc claffy, Michael Kallitsis, Eduard Glatz, and Xenofontas Dimitropoulos. Estimating Internet Address Space Usage through Passive Measurements. *SIGCOMM Comput. Commun. Rev.*, 44(1):42–49, December 2014. ISSN 0146-4833. doi: 10.1145/2567561.2567568.

[8] Fei Du, Yongzheng Zhang, Xiuguo Bao, and Boyuan Liu. FENet: Roles Classification of IP Addresses Using Connection Patterns. In *2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT)*, pages 158–164, 2019. doi: 10.1109/INFOCT.2019.8711412.

[9] Vaibhav Bajpai and Jürgen Schönwälder. A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *IEEE Communications Surveys Tutorials*, 17(3):1313–1341, 2015. doi: 10.1109/COMST.2015.2418435.

[10] Speedtest by Ookla - The Global Broadband Speed Test. URL `https://speedtest.net`. Accessed: 2021-06-18.

[11] Visualizations - M-Lab. URL `https://www.measurementlab.net/visualizations/`. Accessed: 2021-06-18.

[12] University of Oregon Route Views Archive Project . URL `http://routeviews.org/`. Accessed: 2021-06-11.

[13] Measurement Lab. The M-Lab NDT Data Set, (2009-02-11 – 2015-12-21). URL `https://measurementlab.net/tests/ndt`. Accessed: 28-4-2021.

[14] Hari Balakrishnan, Venkata N. Padmanabhan, and Randy H. Katz. The Effects of Asymmetry on TCP Performance. *Mobile Networks and Applications*, 4 (3):219–241, 1999. doi: 10.1023/a:1019155000496.

[15] Lorenzo Colitti, Steinar H. Gunderson, Erik Kline, and Tiziana Refice. Evaluating IPv6 Adoption in the Internet. In *Passive and Active Measurement*, pages 141–150. Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-12334-4_15.

[16] Jakub Czyz, Mark Allman, Jing Zhang, Scott Iekel-Johnson, Eric Osterweil, and Michael Bailey. Measuring IPv6 adoption. In *Proceedings of the 2014 ACM conference on SIGCOMM*. ACM, August 2014. doi: 10.1145/2619239.2626295.

[17] CAIDA AS Rank: A ranking of the largest Autonomous Systems (AS) in the Internet. URL `https://asrank.caida.org/`. Accessed: 2021-06-17.