École polytechnique de Louvain

Internet Performance Toward Open Access Content

Author: **Manon JACQUEMIN**
Supervisor: **Cristel PELSSER**
Readers: **Tom BARBETTE, Auguste BURLATS**
Academic year 2023–2024
Master [120] in Computer Science

Abstract

Open access content is scientific and research content that is freely accessible. It is important to be able to access it because it enables knowledge to be disseminated to a wide audience, particularly in developing countries. The aim of this master's thesis is to measure access to different open access directories on a global scale, from research and commercial networks in IPv4 and IPv6, in order to detect inequalities and access problems.

To achieve this objective, measurements were taken from RIPE Atlas probes in research and commercial networks from all regions of the world. Pings were made 12 times a day and traceroutes one time a day during the whole month of January 2024. The results were analysed by continent, then the situation in several countries was studied in detail. The RTT, the number of hops and the AS path were studied for these measurements.

The measurements showed inequalities between continents, with Africa and Oceania performing the worst and Europe the best. There are also inequalities between research and commercial networks, neither of which is systematically better than the other, depending on the continent. Performance also varies within the same region depending on the repositories, with those hosted by CDNs performing much better.

Acknowledgements

I would sincerely like to thank my promoter, Professor Pelsser, for the time she devoted to me, her precious advices and continual support throughout the year.

I would like to thank Professor Barbette and Mr Burlats for their proofreading.

I am grateful to my parents for their support and encouragement throughout my studies.

Finally, I would like to express my gratitude to my cat Pluche, for her presence and moral support during this period.

Contents

Chapter 1

Introduction

Innovation and the development of new knowledge are based on existing scientific research. However, this content is expensive and not accessible to everyone, particularly in developing countries. With the aim of making this knowledge accessible to everyone, open access was born.

Open access content refers to scientific and academic research that is freely available. It facilitates the sharing of knowledge, helping to improve the quality of research, and enables more funds to be invested directly in research. It is therefore beneficial to everyone, but especially to developing countries, which often lack the budget to access the conventional scientific journals [42]. There are several ways of publishing in open access, the green and gold way.

The green way is to publish its research in open access directories, while the gold way is to pay money to publish in open access journals. This amount is often too high compared to the research budget of developing countries [65].

In addition to these difficulties, some regions may experience poor internet performance to access the repository, making these open access materials even more difficult to access.

It is important to evaluate the accessibility of this content, since it could highlight undiscovered accessibility problems and inequalities in performance to access new knowledge.

Several studies have already been carried out using RIPE Atlas probes to compare performance in different parts of the world [27, 35, 29, 8] showing some disparities between the countries. RIPE Atlas is a tool that enables near real-time measurements from probes deployed in different parts of the world in several types of networks.

The goal of this master's thesis is to investigate the connectivity of open access content on a global scale. This entails conducting internet measurements across diverse network types to analyse the performance metrics such as the latency and the paths employed to access open content repositories.

In this study, we determine whether there are inequalities between continents, between research and commercial networks and between measurements made with IPv4 and IPv6 by studying the Round-Trip-Time (RTT) to open access directories. We also analyse the routes taken by packets in studying traceroutes. To achieve this, measurements are performed using RIPE Atlas probes.

We begin by taking an overview of connectivity worldwide and on each continent. We then present the National Research and Education Network (NREN) and the RIPE Atlas probes. We then look at open access and related works on the subjects. We then present the open access repositories. After, we detail how the probes were selected. We describe the method used to measure and analyse pings and traceroutes. The chapter 4 presents the measurement results. Firstly, on a global scale, comparing continents, then we look at several countries: South Africa, China, India, Argentina and New Zealand. Finally, we propose recommendations for each continent.

Chapter 2

Context

2.1 Background

2.1.1 Internet global situation

In 2023, the number of Internet users worldwide is estimated at 5.4 billion people by the International Telecommunication Union (ITU), this represents 67% of the world's population [57]. However, there are major inequalities between continents and countries in terms of access and performance. 96% of people without Internet access live in developing countries [56].

Some regions of the world, such as South Asia, Latin America and Africa, are about ten years behind the other regions in terms of Internet development[4].

Nevertheless, these regions have a higher growth rate than other regions, particularly in terms of the number of Internet Exchange Point (IXP) [70] or investments in submarine fibre cables [12].

Despite this, Europe has more than a third of active IXP, Asia-Pacific and North America more than 20% each, while Latin America has 10% and Africa only 7% [69]. Nevertheless, it is important that as many countries and regions as possible have their own IXP in order to avoid long detours when sending packets and to guarantee the lowest possible latency [40].

Underwater fibre-optic cables are essential to the Internet, since 99% of Internet traffic passes through them. There are more than 400 of them, covering 1.3 million kilometres. Their advantages are that they are faster, more robust, can carry more data and are cheaper than satellite connections [24, 59].

Internet access and performance are very important because they correlate strongly with the United Nation (UN) and ITU's economic development index [3].

2.1.2 Africa

While 37% of the African population uses the internet, some regions, such as sub-Saharan Africa have even lower internet penetration rates around 28% [2]. Many people could have access to it but don't, however 30% of the rural population have no access to the mobile network, which is one of the main means of accessing the internet [56]. Most of the African infrastructure, such as the cache and cloud infrastructure, is divided between a few countries such as South Africa, and Egypt. [48, 83] and many African websites are hosted outside the continent [37].

Africa has seen the biggest growth in active capacity between 2007 and 2012, after years of a lack of investment. The installation of several cables on the 2 coasts of Africa, including one between Angola and Brazil, is proof of this. However, the land cables at the landing points of the submarine cables are sometimes neglected, which cancels out the benefits of the cables [12]. This result in bad performance both in latency and bandwidth [22]. This is why new terrestrial fibre optic cables need to be installed in many African countries. These installations will make the submarine cables surrounding the continent less essential, but also help to make them more efficient [24].

There is a real lack of connections between countries within the continent itself [2]. As shown by Chavula et al. in 2015 [21], 75% of traffic between African universities goes through Europe. It is more efficient in terms of latency to use Europe or North America networks than the African networks for intra-African traffic [39, 22, 35], geographical distance is therefore not positively correlated with latency in this case. This results in a very high average RTT of more than 150 ms [36] or 200 ms to the most popular websites [22]. The second consequence is that African internet companies have to pay hundreds of millions to non-African companies to transfer internet packets between African countries. The price for international bandwidth is therefore the highest in the world and has a significant impact on the prices of internet providers [2].

However, Africa hosts one of the most important points for global connectivity, the Suez Canal, which is the most important crossing point for Internet traffic between Europe, Asia and Africa [12].

2.1.3 Asia

Approximately 72% of the population in the Asia-Pacific region (APAC) uses the internet [25]. Asia is divided into different regions when it comes to Internet development, with some regions such as East Asia far ahead technologically, unlike South Asia, where development is lagging behind [76, 27]. However, it is the fastest-growing of the developing regions [4]. They can also take advantage of the infrastructure in the east of the continent for certain routes and have direct and faster routes to the United States, like Malaysia for example which, can pass through Japan and Singapore [3].

Asia is the anchor point for many important submarine cables, including those to Europe via the Suez Canal and transatlantic cables. Two third of Chinese submarine cables are directed towards the USA, despite China's desire for Internet autarky [12].

Some regions of Asia, such as the Indian subcontinent and the Middle East, have seen a huge increase in investment in submarine cables, but this has not been followed by investment in terrestrial cables, so as in Africa, the advantages have not been fully realized [12]. However, the installation of submarine cables in the South East Asian region has greatly improved the RTT and throughput [3]. This installation has had an impact not only on performance but also on the country's overall economy, for example, Indonesia's GDP per capita rose by 5.4% between 2012 and 2017 [67].

However, submarine cables in Asia are more prone to failure due to natural causes, which can have a major impact on internet traffic [24]. To avoid such problems, there is need of reliability like in Japan during the 2011 earthquake[32].

2.1.4 Europe

Over 92% of European homes have an internet connection. Europe is the continent with the most network infrastructure. It has the highest number of active RIPE Atlas probes, the highest number of Autonomous Systems (ASes), IXP and lots of data centres. Each European country has at least 7 data centres and an IXP, which means that each country can manage its own traffic without the intervention of another [24].

The continent is well connected to the rest of the world via 250 fibre cables, 2/3 of which are underwater and 1/3 terrestrial with neighbouring non-European countries. Europe and North America have the largest international capacity. There are also numerous connections between the Middle East and North Africa, as well as to Asia. Connections from the rest of the world are less direct, in particularly South America and Oceania, where the traffic passes through North America.

2.1.5 North America

The number of Internet users in North America was estimated at 92% of the population in 2023 [25].

The United States plays a major role in the global Internet as it hosts all the GAFAMs (Google, Apple, Facebook, Amazon and Microsoft) and major Internet companies, data centres, root servers and domain names widely used as .com or

.org. They are a hub for global connectivity, with 80% of Internet flow passing through them [51].

2.1.6 Oceania

Oceania is very dependent on submarine cables for Internet traffic since the continent is mainly composed of islands, which complicates the deployment of infrastructure [20].

There is a disparity between Australia and New Zealand, which have good infrastructures and highly developed links with the USA, and the smaller countries, which are less equipped and connected. Australia is therefore funding connection setup between the smallest islands and the largest countries, such as the USA and Japan, to improve redundancy and reliability [10].

2.1.7 South America

Internet penetration in Latin America is estimated at 70% [25]. As explained earlier, South America has a high growth rate in terms of infrastructure, with an increase in the number of IXPs [70]. This is partly due to 55% of actions taken by the various continent countries' governments by creating laws, federal or non-profitable agencies, providing funds [16].

South America is strongly connected to North America and partly dependent for the traffic to other regions of the world. In 2015, around 80% of internet communications between Latin America and Europe passed through the USA [51]. Some Latin American countries, such as Brazil, wanted to be independent from USA since the data spying scandals by creating their own routes via submarine cables[14]. This resulted in the creation of BELLA in collaboration with GÉANT and RedCLARA [66]. In 2018, to improve latency between Africa and South America, the first south-to-south cable undersea cable, between Angola and Brazil, was deployed. This reduced latency between the two continents, but some packets sent from Europe, Asia and Africa towards Africa increased their latency because of unnecessary detours [38].

2.1.8 NREN

Each country has its own National Research and Education Network (NREN). Its aim is to develop communications infrastructures and services dedicated to research and education institutions. One of the most Internet2 [52], also known as UCAID in the United States, brings together more than 200 universities and commercial partners, with the aim of developing IPv6 and other Internet technologies such as multicast. In China, CERNET [19], the China Education and Research Network, and CERNET2, its IPv6 equivalent, bring together more than 3,000 entities [62].

In South Africa, SANReN [77] and TENET [84] are cooperating to offer the South African NREN. The South African National Research Network, provides the network infrastructure for research and education, while TENET, The Tertiary Education and Research Network of South Africa, manages the network services.

In Europe, we can mention BELNET in Belgium, RENATER in France, JANET in Great Britain. These European countries are gathered in a single organization called GÉANT [41], with a common goal to develop the European Internet research network.

Like the NRENs of Europe, the NRENs of other continents are generally brought together in organizations to facilitate connectivity between the regions of the world. One example is UbuntuNet Alliance [86], which brings together NRENs from 22 countries in Eastern and Southern Africa. They are partners with, among others, WACREN, the West and Central African Research and Education Network, ASREN, the Arab States Research and Education Network, and RedCLARA, the Latin American Advanced Networks Cooperation. AARNet, Australian Academic and Research Network and APAN, Asia Pacific Advanced Network can also be cited.

These organizations are taking initiatives among themselves to create global collaborations, such as Networks for European (NEA3R) [64], American, African, and Arctic Research, and Orient plus [68], which links CERNET and GÉANT.These cooperations provide the best performance, resilience and accessibility for research networks.

2.1.9 RIPE Atlas probes

RIPE Atlas is a platform for Internet measurement worldwide in real time. It is made up of thousands of users who host measuring devices in all types of networks, particularly in commercial and research networks. This network allows the study of Internet in all its glory and extract data from it. In addition to the data already available, users can take personalized measurement from all probes to the destinations of their choice. RIPE Atlas offers the possibility of taking measurement from 2 types of devices. The first and most deployed are the probes. These are small hardware devices connected to a router, they use little bandwidth and cannot retrieve information on the host's content. The second type of device are anchors, which are probes with greater measurement capabilities. There have been 5 generations of probes deployed over the years. It is recommended to avoid versions 1 and 2 as they have a higher latency than version 3 [7] Measurements sometimes suffer interference when they are made at the same time [46]. They are used as probes to assess the connectivity and reachability of a particular region.

These probes can ping, traceroute, SSL/TLS, DNS, NTP and HTTP to specified destination.

The number of probes continously increased since the platform was created, and by 2024 there were more than 12,000 probes and more than 800 anchors connected in real time. On May 4, 2024, a total of 17,102 connected and disconnected (excluding abandoned) probes were registered on the site. All these probes are distributed around the world, but there is a huge geolocation bias [79]. As shown by Sermpezis et al., the distribution of probes is non-uniform in several Internet measurement platforms. The authors propose a calculation to measure the bias and show that RIPE Atlas has a lower bias than RIPE RIS and Routeviews. However, the number of probes on May 4, 2024, shows this bias. Almost 60 % of probes are implemented in Europe. Asia and North America have a little over 15 % of probes each, but probes are unevenly distributed, with a few countries with many probes and many countries with very few probes. For example, the United States alone had 2,386 probes which represent 13.77 % of the total number of probes. Africa, Asia and Latin America all together have 2 percent of probes deployed.

Continent	number of probes	percentage
Africa	381	2.24
Antarctica	1	0.01
Asia	2,859	16.45
Europe	10,062	59.24
North America	2,979	17.18
Oceania	475	2.74
South America	416	2.42

The United States is the country with the most probes installed, while the other 4 countries with the most probes are all in Europe: Germany, France, Great Britain and the Netherlands. The rest of the top ten is made of two other European countries, two Asian countries and one North American country. The number of probes for all countries can be found in the appendix.

Country	number of probes	percentage
US	2,386	13.77
DE	2,234	12.89
FR	1,346	7.77
GB	809	4.67
NL	807	4.66
RU	734	4.24
IT	542	3.13
CZ	400	2.31
CA	394	2.27
JP	347	2.00

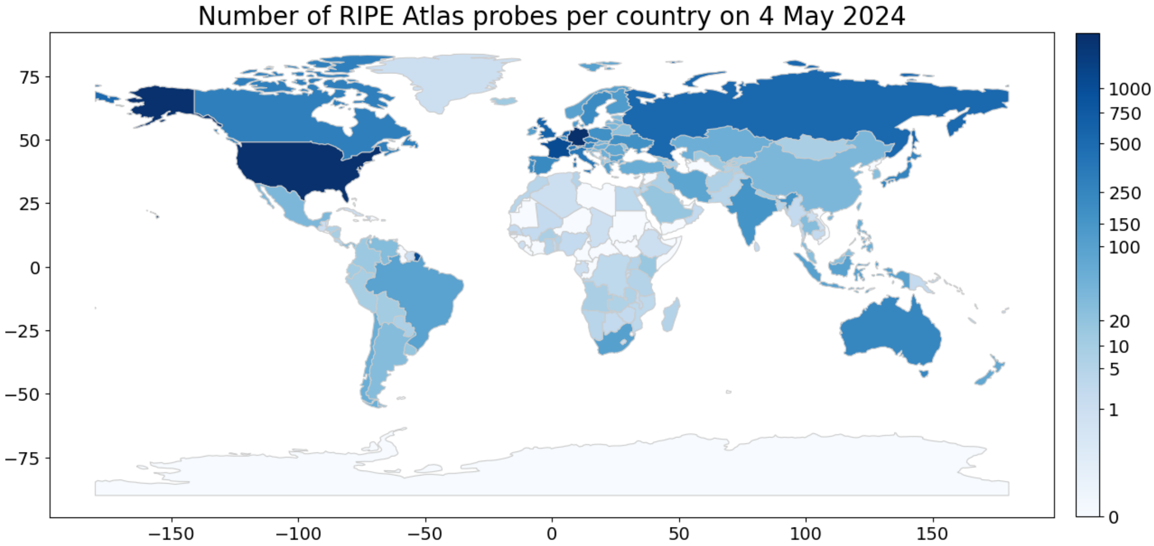

Figure 2.1: Number of RIPE Atlas probes connected by country on 4 May 2024

The number of inhabitants is not balanced with the distribution of probes. When examining the countries with the most RIPE Atlas probes, they cover a small proportion of the global population. The 10 countries with the most RIPE Atlas probes cover only 11% [1] of the world's population, but more than 58% of the total number of probes (excluding abandoned ones).

Conversely, if we take the 10 most populous countries, which represents 57%[2] of the total population of the globe, only 23% of probes (excluding those abandoned), are deployed. This shows that the majority of the world's population is very partially covered by probes, and that certain countries are over-represented.

The coverage of Internet users in comparison with the number of IPv4 probes, i.e. the percentage of users who are in a network that has at least one atlas probe, also shows a geographical disparity, particularly between developed and developing countries [73]. Europe, North America and Oceania all have very high coverage, at over 85%, while Latin America and South Asia have coverage of around 30%. In Africa, some countries have more than 50% coverage, but the majority are at 0%. For IPv6, global coverage is much lower for all regions, but some European countries are maintaining very high coverage, close to 90% [74].

Another type of bias exists: topology bias. It consists in having a bias on large networks such as Internet eXchangePoints (IXP), which are more likely to install a lot of probes to monitor their network metrics [79].

[1]The percentage of population has been calculated using data from UNFPA `https://www.unfpa.org/fr/data/world-population-dashboard` on 27 May 2024.

[2]See footnote 1.

2.1.10 Open access

Scientific content is one of the most important foundations of scientific innovation and learning. It is necessary for research, since new knowledge is enabled by on old knowledge. This includes different types of medium: journals, conference papers, books, databases, etc.

One way to accessing and distributing this content is through libraries, where content can be consulted or borrowed. However, they are often difficult to access, especially in developing countries like Africa [1] or India [42]. The Southern African Universities association (SARUA) conducted a study in 8 southern African universities, interviewing 76 university members. 54 % of respondents thought that the region's research results were not accessible. Most of the other respondents think that they are not easily accessible due to numerous obstacles. The lack of deposit and publication due to the difficulty of publishing their research leads to a lack of awareness of what has already been produced, within their own institution or region. Another obstacle is the competition between researchers for funding, which is very limited, making some researchers reluctant to share their results.

The lack of funding for African universities is a problem for the acquisition of new content, due to the high cost of subscriptions.

Journal prices vary enormously depending on a whole range of factors: the publisher, the revenue, the rate of publication, the negotiation, the size of the university, the packages of several journals, whether it includes the possibility of publishing in the journal. In 2009, American universities were paying between 1 and 2 million dollars a year to subscribe to the Elsevier Freedom package, which includes over 2,000 journals. Others were paying between 200,000 and 500,000 dollars for the Springer package [11]. Ten years later, the University of California was paying 11 million dollars a year for its Elsevier subscriptions. In all, it was paying more than 40 million dollars in subscription fees for journals [44]. The University of Lorraine paid more than 40,000 euros to IEEE and more than 700,000 euros to Elsevier [88]. Nowadays, universities and institutes in France pay around 3 million euros a year in subscriptions to various journals [13, 60].

Open access could be a solution to this problem. Open access (OA) provides free access to scientific content for the public [82]. OA has many advantages, increasing access and dissemination of scientific resources, and therefore their visibility and potential impact. Sharing knowledge promotes economic, social and innovation development. They increase the quality, speed and efficiency of research. They also open up knowledge in a given field to other disciplines, but also to other communities. However, care must be taken with the quality of some non-peer-reviewed content.

Furthermore, scientific research is largely financed by public funds [89, 87, 90, 47]. For example, programs by the European Union [34] such as Horizon 2020, NSF in Us, NSFC in China, or regional and national subsidies fund research programs [90]. It is therefore fair that the results should be made public and accessible so that the public can benefit from them.

The Internet is the most efficient way of communicating and disseminating free open access content, thanks to open access directories. The first of these was arXiv in 1991, initiated by physicist Paul Ginsparg. It is the first site to host preprints and pre-reviews. Since then, many repositories have been created in all scientific field, providing easy access to this content.

There are two ways to publish in open access: the green and gold way. The green ways consist in publishing its paper by its own in a repository, it is generally faster than the second way. The gold way consists in publishing in open access journals but which requires financial compensation that is often very high and difficult to pay for researchers in the poorest countries [65]. There is a third way, less used, which is a hybrid of the two first. The author pays a fee to publish one open access article in a subscription journal.

2.2 Related works

Numerous studies have been carried out to examine performance from different points of view. The first is to compare IPv4 and IPv6.

First, Forough et al. [43] studied the stability of the paths taken by IPv4 and IPv6. Their measurements were performed using the NORNET CORE network, which includes university and research centre sites in Europe and China. These research networks have the advantage of all having IPv6, but some Internet Service Providers (ISPs) have not deployed it. They discovered that the paths taken by IPv6 are less stable, i.e. they change more over the course of a day. They also saw more load balancing in IPv6 and did not observe any significant difference between the length of IPv4 and IPv6 paths.

Stephen D. Strowes in [81] has studied peak cycles in IPv6 measurements to Yahoo over 3 weeks. It identified different patterns between weekday and weekend days for fixed line providers. The IPv6 requests are more frequent in the evening during the week and all day at weekends. Conversely, mobile providers do not report this difference for IPv6. For IPv4, there are more requests during office hours, as the majority of office requests are in IPv4.

Dhamdhere et al. [31] have analysed old Border Gateway Protocol (BGP) records and new measurements to analyse IPv6 evolution trends. Their conclusion is that the IPv6 network is growing, but not uniformly. In terms of topology, IPv6 is deployed in the core network and not at the edges. Carisimo et al. [17] study in detail the adoption of IPv4 and IPv6 in the core network of several Content Delivery Netowrk (CDN), which have a very important role in the global adoption of IPv6.

Dhamdhere [31], Stephen D. Strowes [81] et Czyz et al. [28] observed a disparity in the adoption of IPv6 depending on the geolocation, with the majority in Europe, North America and the Asia-Pacific region.

Li [61] and Dhamdhere [31] conclude that for similar AS paths, IPv4 and IPv6 performances are similar but that if the AS paths are different, IPv6 is worse than IPv4.

Other studies have looked at the performance of cloud providers on a global scale. [30] shows that performance in terms of latency is similar for developed countries, regardless of the provider, thanks to the significant development of network infrastructure and data centres. Performance in developing countries is worse and depends a lot on the distance between the servers and the user. Corneo [27] has also made this observation. For example, Europe and the USA have many cloud data centres and therefore have a very low RTT and number of hop to access the cloud. On the other hand, users in Africa experience much higher latency when accessing the cloud, and in some cases are forced to go to North America or Europe. Asia shows unequal latency between its different regions, depending on whether or not there is a close data centre. The authors therefore conclude that geographical distance from the nearest server has a real impact on cloud access performance, and they recommend deploying more data centres in regions that do not have one in order to improve performance.

Chapter 3

Methodology

3.1 Open access repositories

This section presents the methodology employed to select the open access repositories for which performance measurements were conducted, as well as the approach used to deduce their hosting details. The last aspect involves identifying whether the hosting provider is a CDN or not.

The open access sites selected are 15 in total, although performance has been evaluated for 14 of them, as two sites share the same IP address with one site being hosted by the other. These sites cover all scientific fields and includes several millions of articles and preprints. The total numbers of references including preprints, pre-review articles, published and reviewed articles available on each repository on the 4th April 2024 is shown in the table 3.1. Some article may be included in multiple repositories, for example, articles published with Hal can also be published automatically with arXiv on author's demand [45].

They are located only in Europe and in the United States. To our knowledge, there are no sites hosting open access content on other continents.

To identify the hosts, we first performed DNS lookup to find the IP addresses of the open access sites. The host names were then deduced from the field *server* from the https responses and with geolocation databases such as MaxMind and IpInfo based on the IP addresses found with the DNS lookup. To obtain the most accurate locations of the site hosts, the database has been used.

CDN consists of multiple caches deployed all over the world by content delivery provider to optimize the quality of the web browsing by reducing the loading time of web pages [78]. When a user wants to access content hosted by a CDN, a request it sent to the closest replica offering the user the best performances [15]. It is also

Open repositories	Number of references
Open Alex	240,000,000
sci-hub.st	88,343,822
sci-hub.se	88,343,822
Europe PMC	43,868,772
HAL	3,629,652
arXiv	2,454,476
Research Square	295,918
bioRxiv	230,032
OSF Preprints	150,016 (between 2016-2024)
MDPI Preprints	64,167
Authorea Preprint Repository	46,936
medRxiv	42,437
Cambridge Open Engage	25,370
OAK	6,519
Africarxiv	+600

Table 3.1: Number of references of each open source repositories on 4th April 2024

useful for reduce bandwidth and taking into account the network's condition to perform load balancing [58].

To determine whether the host is a CDN, we have followed the method proposed by Saverimoutou and al. [78]. The principle is to analyse the response headers of HTTPS requests, which are obtained by using 'curl -I destinations' from the command tool. The presence of *cache-hit* in the response header indicated that the content is delivered by a CDN or a cache node. CDN can use different terms to describe a cache-hit: it's "CF-Cache-Status" for Cloudflare, it's "cdn_cache_status" for Google Cloud and "X-Cache" for Fastly and Amazon [63]. Amazon's services are also identifiable through "CloudFront". To confirm our observations, we used CDNFinder [18] which is a database that identifies CDN from an IP address or URL.

This method has enabled us to identify the following 15 directories presented in table 7.2, their host, their IPv4 and IPv6 addresses and their respective locations. Of these directories, 11 are hosted in a CDN: Open Alex, MDPI Preprints, Research Square, medRxiv, Authorea Preprint Repository and Cambridge Open are hosted by Cloudflare; arXiv is hosted by Fastly ; OSF Preprints and Africarxiv use Google cloud and sci-hub.st and sci-hub.se use Ddos-guard. 10 are hosted by American companies and 5 are hosted in Europe.

The sites hosted by Cloudflare all use anycast, which means that the same IP address is advertised from servers located all over the world, with traffic redirected to the data centre closest to the user. This reduces RTT, improves network resilience and reduces the risk of Distributed Denial-of-Service (DDoS) attacks [71]. Cloudflare has many data centres spread across every continent, which means we can get good results from many parts of the world [26]. One way of identifying whether an IP address is anycasting is to analyse the traceroutes and look at the last hops before the one that indicates the final IP address [71]. It is possible to identify the location of previous IP addresses using the name of the city, airport or country in the name associated with the IP address. If, from several different locations, the last hop before the destination is in different countries but leads to the same IP address, it is highly probable that the IP address is anycasted. One way of checking these assumptions is to use databases that group together anycasted IP addresses, such as anycasted.io [5].

All these directories have an IPv4 address, but only 4 of them are accessible via an IPv6 address. They are all hosted by Cloudflare, but 2 of those hosted by Cloudflare do not have an IPv6 address. All IP addresses have been obtained through DNS look up and confirmed by RIPE Atlas measurement.

Between November 2023, when the first observations were made, and January 2024, when the measurements were taken, several sites changed their hosts: arXiv has moved from local hosting at Cornell University to the CDN Cloudflare. Similarly, Authorea Preprint Repository decided to migrate from Amazon to Cloudflare. This has obviously changed their IPv4 address but has not affected the country in which they are hosted.

3.2 Probes selection

This section presents the methodology used to select the probes from which the measurements to the open source directories are performed.

The first step is to identify the Autonomous System Numbers (ASNs) from educational and research organism and from commercial affiliations, i.e. Internet Service Providers (ISP), cable providers, Digital Subscriber Line (DSL) operators and Network Service Providers (NSP) by questioning the database of Peering DB. This is a database that regroup networks, its aim is to facilitate interconnections between networks. In Peering DB that correspond to the filter NSP, Cable/DSL/ISP and Educational/Research. The choice was made to restrict ourselves to these types of networks, since the people most likely to visit repositories with open access content use mostly research networks in university for example or in commercial networks, i.e. those that are accessible to the public. Networks type that are not included but registered in Peering DB are Not Disclosed, Content, Entreprise, Non-Profit, Route Server, Network Services, Route Collector and Government.

The second step is to obtain all the ASNs in which there is at least one connected probe within RIPE Atlas, excluding abandoned and disconnected probes. RIPE Atlas is a tool enabling the measurement of pings and traceroute to user-selected sites from probes spread around the world with numerous customizable parameters.

Subsequently, the list of ASNs connected in RIPE Atlas from research and commercial networks available in RIPE Atlas is fetched by comparing the ASN present in the precedent steps.

All the information about the probes present in these ANSs have been recorded during the month of November 2023, collecting the data once a day. This process aims to assist us to select the most stable probes for our measurements. To determine their stability, the frequency of probe connections have been calculated, taking into account the number of times a probe has been connected during the period, divided by the total number of days. Then, the probes have been sorted in ascending order by country.

Once the list of the most stable have been created, one probe from both research and commercial networks have been selected for various countries worldwide, the probes support either IPv4 or IPv6 or both, spread across all continents. When this was possible, preference was given to probes that show the highest frequency of connections.

3.3 Measurement

The measurement has been performed using RIPE Atlas, the aim of these measurements is to measure the RTT, the AS path and the number of hops between the selected sources and the open access sites. To do this, ping and trace routes have been performed at different times.

The first measurement campaign involved trace routes being sent 3 times a day for 3 days in December 2023. Three packets were sent for each traceroute. The goal was to analyse the different AS paths used and determine whether any of them were similar, thereby reducing the number of source probes.

The second measurements campaign consisted of taking measurements over a one-month period. Pings were sent every two hours from midnight Brussels time, with one packet per ping, and trace routes were sent once a day at midnight for the whole of January 2024. In total, there are 339,090 measurements : 312,983 pings and 26,107 trace routes.

3.4 Measurement analysis

3.4.1 Ping

The results obtained with the pings have been analysed from 3 different angles. The first combines all the measurements, regardless of the day or time, without distinguishing between commercial and search networks for all the countries on a continent, to obtain the overall average RTT for that continent. The second average was calculated in the same way, but with a distinction between research and commercial networks, so that measurements from commercial and research probes could be compared within the same continent.

The second way of analysing ping results is from a daily perspective. The aim of this analysis is to study congestion on a typical day in January 2024. To do this, all the pings measured at the same hours were added together and then divided by the number of pings measured for that hour, to obtain the average. The graph resulting from this analysis shows the local times. The local time was calculated using the latitude and longitude recorded in the information from the RIPE Atlas probe, then the time zone was obtained from these geographical coordinates using Google data.

The last analysis angle covers the whole month of January 2024. To do this, the average of all the measurements taken each day, regardless of the time of day, has been calculated for each day of the month. This analysis makes it possible to study the evolution of RTT over the whole month in order to detect the possible presence of congestion on specific days.

To analyse in more detail the evolution over January and a typical day, peaks were detected on the basis of a threshold. To do this, we use the standard deviation of degree of freedom of 2 and the mean. For each element of the day or month, if the measured value exceeds the mean by more than one standard deviation, then a peak is detected and is represented by a red dot on the graph.

3.4.2 Traceroute

The first step in traceroute analysis is to deduce the AS path from the IP supplied in the result of the traceroute. To do this, we used the RIPE Stat tool. If any IP information is missing, we put an "x" in the AS path. When we have an IP address, but RIPE Stat has no information on it, we put "no info" in the AS path. This happens when the IP address is reserved, i.e. it can be private or multicast.

Once the AS path has been obtained, we look at path changes. To do this, we take the AS path, remove the non-useful information, x and no info, and remove the AS that are repeated to detect radical changes of AS. We then obtain a list of all the ASes that have been visited at least once. We can then compare it with the next AS path, and if it is different, we indicate this with a red dot on the graph.

We don't take all changes into account because the AS path changes very regularly without changing the path taken. For example, the AS path prepending consists in announcing its AS several times in the AS path to influence the routing of packets, which changes the AS path but not the path taken. Another example could be an IP address that is not available in one AS path but becomes available in another AS path. These changes are generally of little consequence in our case, since the visited ASes and their locations remain the same.

The next step is to identify the countries and cities through which the AS paths transit. To find the most precise geolocation possible for an AS, we look at the response from the DNS lookup provided in RIPE Atlas. These responses can provide a lot of information about the location of the server through which the packet is passing, but they can also provide no information apart from the address IP or the AS number, for example in the case of a CDN that uses anycast.

The DNS lookup response may provide an abbreviation or the name of the city like "par" for Paris, the code of an airport, such as "JFK", or a country, such as "za" for South Africa, "fr" for France, etc. These codes may vary for the same city depending on the AS managers. For example, "hnk" and "hkg" both indicate Hong Kong, the first at Tweelve99, the other at Cloudflare. Some companies, such as twelve99, provide a list of available routers together with their name and city, so it is possible to find out from the DNS lookup response which specific city is being used to transfer the packet [85]. To check the information, we can couple the information with RIPE Stat which provides the locations for certain IP and check the networks on the official websites of the organizations if the paths deduced are possible. For example, on the orient plus website, we can see that between China and Europe, one of the European anchor points is in London, which corresponds to the information on the look-up DNS which indicates "lon".

By combining all these sources, it is possible to approximate the path taken by the traceroute.

The final analysis compares the last RTT and the number of hops provided in the traceroute to see if there is any correlation between the two. The number of hops is retrieved from the response of a RIPE Atlas traceroute.

Chapter 4

Result

4.1 Global observation

Of the 14 directories pinged and tracerouted, all are accessible via an IPv4 address and only 4 via an IPv6 address: docs.openalex.org, medrxiv.org, authorea.com, www.researchsquare.com, all hosted by Cloudflare. This can be explained by the fact that Cloudflare has a policy of promoting the use of IPv6 by offering free IPv6 support for its customers [75]. Despite this, two of the 6 sites hosted by Cloudflare, Mdpi and Cambridge, have apparently chosen to have only one IPv4 address.

Of the 14 sites probed in IPv4, for 3 of them : Sci-hub.st, Sci-hub.se and BioRxiv, the pings gave no results, none of the probes reached their target. However, thanks to traceroute, it is possible to approximate the RTT and AS path used despite the fact that the probe did not reach its final destination. In fact, the traceroutes reached a location close to the one of the site's hosting, without actually reaching it. For example, for the two Sci-hub sites, the last IP address reached by the traceroute is in Amsterdam in the Netherlands, the same city where the site is hosted (according to IP info). We can assume that these are the last steps before the site servers, but that the sites are protected by a firewall or other security device that prevents pings and traceroutes.

Since IPv6 is much less widely deployed and the results can be very different from IPv4, it is important to analyse the two separately and then compare them. We started by analysing the ping results. We began by analysing the results of measurements made with RIPE Atlas during January 2024 in IPv4. The average RTT for each continent was calculated without distinguishing between research and commercial networks, in order to compare performance between continents and identify the regions with the best and worst results.

In the next step, we separated the results between the research networks and the commercial networks in order to identify whether there were differences in performance and to establish which was the most effective for each continent. Then, we did the same process for IPv6 and finished by comparing the results between IPv4 and IPv6 to see if the 2 protocols had similar performance.

In the second part we analysed some specific countries: South Africa, China, India, New Zealand and Argentina. For each country, the average RTT over a typical day in January was studied using measurements taken every 2 hours to identify peaks in daily use at certain period.

We also looked at the average RTT for each day of January to see if there was congestion on one or more days and the paths taken by the AS path to understand the country specific results, first in IPv4 and then in IPv6.

4.1.1 Connectivity of probes

When making measurements using RIPE Atlas, there is no guarantee that the desired measurement results will be obtained. There are several reasons for this, and the probes may be disconnected at the time of measurement for various reasons, another reason is that requests do not always reach their target, either because of a problem with the probe or the server. It is also possible that there are no reports available from RIPE Atlas. All this means that many of the measures requested were not carried out or did not achieve any results. For example, of the 50 IPv4 probes requested for pinging during each measurement campaign, 3 did not provide any results during the month of January because they did not take part in any measurements. 3 probes gave no results for one day, and one gave no results for 4 days (2 times 2 days in a row), which probably represents a failure or temporary disconnection. One probe had more serious problems, as it only gave results on the last 2 days of the month.

For IPv6 out of the 48 probes, 8 probes gave no results at all, one of them was the Japanese research network probe which gave also no results in IPv4. It is important to note that we have no results for probes in commercial networks in North America, as all three gave no results for the whole month.

In addition to these 8 probes, 3 have unreachable status, i.e. there is a packet loss of 100% for pings, but approximations have been obtained for traceroutes. In addition, 3 other probes gave no results on certain days in January, one for 2 days, one for 4 days and one gave no results from 14 January onwards, suggesting a problem with the probe at that time.

This unavailability must therefore be taken into account when studying the overall results.

4.1.2 Analysis by continent

IPv4

The average RTT measured throughout the month of January 2024 to the 11 repositories for which the pings worked varies greatly between continents towards the same site, but it also varies for the same continent towards different sites, as shown in figure 7.1 which represent the mean RTT for each continent during January 2024.

If we look at the average, Europe has the best time, less than 60 ms. The two American continents and Asia have similar results, at around 100ms, while Africa is at 140ms and Oceania at 160ms.

When we look at the results site by site, as shown in figure 7.2, we can see that Oceania has the worst RTT in 6 out of 11 cases, showing a real connectivity problem, which may nevertheless seem logical given the location of the continent. Africa has the worst performance to 4 sites and Latin America to one site.

However, we can see that South America has the best RTT for measurements taken to 7 of the 11 repositories and Europe for the remaining 4 sites (including the 3 European sites). In Latin America, there is a huge disparity between sites hosted by a CDN and those hosted in Europe. South American probes take less than 40 ms to the USA and 250 ms to the EU on average, this can be explained by the greater connectivity between North and South America via submarine cables than towards Europe. On the contrary, European probes are much more stable, regardless of whether they are hosted.

In general, performance towards European sites is worse than towards US sites. Another clear example is for probes from Oceania, which take twice as long to reach an EU directory as an US one, averaging 250 ms for Europe compared with 120 ms for the US. The reason is that there is no direct submarine cable between Oceania and Europe, so the route has to pass through other continents to reach its target.

Within Europe itself, probes from outside Europe and North America that want to reach Oak.novartis, hosted in Ireland, take longer than sites hosted in the UK or France. Ireland is less connected than France or the UK, which are popular submarine cable access points for intercontinental connections.

The figure 7.2 shows the mean RTT for both research and commercial network for each continent. Research networks perform better on all continents except Africa and Oceania. European commercial and research network probes both have the lowest average RTT for all sites compared to other continents.

We can see in the figure 7.2 that the worst average results are achieved by commercial North American networks in 5 out of 11 cases. This result is surprising given that most of the sites are hosted by CDN which all have servers in the United States, while the RTT for sites in Europe is even higher. In comparison, the North American research probes have the best results in 8 out of 11 cases, which corresponds to all the sites host by a CDN in the USA. This shows that there is a very large disparity between the research and commercial networks in the United States. This is due to the fact that internet2, which connects research institutions and universities, has a very extensive and efficient fibre optic network [53].

In the other cases, it is either the African research probes (4 out of 11) that have the highest RTT or the commercial probes from Oceania, with 2 cases out of 11. The probes in the research networks in Oceania perform a little better than those in the commercial networks, but are still very high compared with the research networks in other continents.

Figure 7.3 shows the average number of hops per continent for each open access directory, excluding the measurements that did not reach their target, which removes 27% of the measurements, if we count the 14 sites including the 3 sites that gave no successful measurements. Without these 3 sites, the number of measures excluded is only 0.0017%. These 3 sites have been excluded from the analyses because, given that the traceroutes were not completed, the hop numbers and the RTT are not representative of the real situation. The number of hops is higher for European probes and lower for Asian probes. The average number of hops varies little according to the sites hosted by a CDN. The number of hop is much higher towards European sites, particularly in UK. Oceania, Africa and Asia have the lowest number of hops and Europe has the highest number.

We can therefore conclude that the RTT does not really have an exact link with the number of hops. Europe has both the lowest RTT for European sites and the highest number of hops. The RTT varies greatly between continents but the number of hops vary much less, for example towards arxiv.org, Asia and Africa have the same number of hop (7.65) but the RTT is more than twice higher for Africa. This is also the case if we compare destination by destination. For example, for all the sites hosted by an American company, Africa takes an average of 8 hops, but the RTT can vary from 60 ms to 175 ms.

If we compare the search and commercial networks, as shown in Figure 7.4, for Asia, Africa and Europe, the number of hops is always higher in the research networks than in the commercial networks, for the 3 other continents it is always the opposite except in few cases.

In conclusion, for IPv4 performance completed during January 2024, Europe has the best average RTT and Oceania and Africa the worst.

Research networks have better results in general, there can be big differences between research and commercial networks within the same continent. Probes generally take longer to reach sites hosted in Europe than those hosted by a CDN. The number of hops varies very little between measurements taken on different continents and has no direct correlation with the RTT.

IPv6

As the figure 7.5 shows performance completed during January 2024 for the 4 sites that support IPv6 varies greatly between continents, as it does for IPv4. For the average of the 4 sites, Asia and Africa have the lowest performance, they are both above 80 ms, Oceania around 70, while Europe is at 40 ms and North and South America both below 30 ms. This might seem relatively logical given that the 4 sites are hosted by Cloudflare, which have a lot of data centres geographically close to these regions. However, it should be noted that for North America, this only records results for research networks.

When we look at each site in detail, performance is fairly similar for the 4 sites, except measurements made from Asia and Africa to Medrxiv, which have a higher average RTT than the other sites. Performance is therefore close to the average for the 4 directories described above.

Performance also varies within the same continent between research and commercial networks, with research networks performing better overall, as shown in figure 7.6. The figure describes the mean RTT of research and commercial network in IPv6 for all continents. Of the 5 continents that have results for the research and commercial network, the research networks have better results, except for Africa and South America.

Looking site by site, we can see that for Africa, research networks are slower for 3 sites, up to more than twice the RTT of commercial networks for 2 of the 3 cases. For Asia, commercial probes are up to 150 ms slower in the worst case than research networks. European commercial probes are also slower than research probes. For South America, however, the opposite is true. Finally, the probes for the Oceania network types have very similar mean RTT for half the sites. In the 2 other cases, the search networks are slightly faster.

The number of traceroutes that did not reach their target was 8% in IPv6. The number of hop for all continent is represented by the figure 7.7. The number of hops to access the 4 open access sites in IPv6 follow the same trend as demonstrated in figure 7.10, which is explained by the fact that the 4 sites are hosted by the same host. Therefore, the AS path follow the same routes, to join the Cloudflare network.

The figure 7.8 shown the mean number of hops from all continents to the 4 sites supporting IPv6 and the mean of the 4 sites combined distinguishing between research and commercial networks. When comparing research and commercial network, in Africa, Asia and Europe, the research networks always have a higher hop. For Oceania and South America, it's the other way round. However, for Europe and South America, the number of hops is very close.

With regard to response time, the same trends can be observed as for IPv4, there is no obvious link between the two.

To conclude, in IPv6, all the measurements were made towards sites hosted by Cloudflare, the average RTT and number of hops are similar for the 4 sites. Europe, America and South Africa have very good RTT in a range, while Africa and Oceania have poorer results. Research networks generally perform better than commercial networks.

Comparison between IPv4 and IPv6

Generally speaking, IPv6 is faster than IPv4, as we can see from the figure 7.9. For example, all RTT measured towards Openalex except for the research network in North America and the commercial network in Asia, are faster with IPv6. For Medrxiv, both types of networks in Africa are faster with IPv4. For Researchsquare, only the probes in commercial network in North America are a little bit faster with IPv4 than with IPv6. For Authora, research networks in Africa, North America and South America are all 3 slower with IPv6. In Europe, IPv6 has always better performances regardless of the type of network.

The figure 7.10 shown the number of hop for the 4 repositories which support IPv4 and IPv6. The number of hops is comparable between IPv4 and IPv6, and the order of scale and trends are almost identical. However, IPv4 seems to be more stable, as only 0.0017% of measurements were unsuccessful if we excluded all the measurement to the 3 repositories that have not been a success, compared with 8.00% for IPv6.

4.2 Case study

The below case study analyses the performance of 5 countries: South Africa, China, India, New Zealand and Argentina. The research and commercial network probes in IPv4 and IPv6 have all given results for these 5 countries, so it is interesting to study them in detail and get an overall idea of the performance obtained in these countries.

4.2.1 South Africa

South Africa is the African country with the largest Gross Domestic Product (GDP), however it suffers from great inequalities, including social economic. It is one of the countries with the most developed internet structures in Africa, owning all the clouds and data centres on the continent except one[83]. It has the highest number of internet users in sub-Saharan Africa, with 38.13 million users in 2021, or almost 60% of the population. However, there is a huge disparity between rural and urban areas, with only 1.7% of rural households connected to the internet [6]. It is by far the African country with the most RIPE Atlas probes, with almost a third of the continent's probes on 4 May 2024 and half of the continent's connected probes. The country is mainly connected to other African countries via submarine cables and terrestrial fibre. It also has 2 submarine cables to Europe and one to Asia. The country has a highly developed network of fibre research roads compared to other African countries [55]. The details of the terrestrial fibres are available in the table 7.3. This network means that the country serves as a communications pole for the East-West Africa [35].

The South African probes selected to take measurements during January 2024 are both different. The pings were made every two hours from midnight Brussels time during 1 month. The traceroutes were carried out every day from midnight in January 2024. The measurements sent a packet. Four separate probes were used to carry out the measurements, two in research networks and two in commercial networks, split equally between IPv4 and IPv6.

Figure 4.1 shows the evolution of the RTT during January 2024 for measurements made from the South African research and commercial probe in IPv4. It can be seen that the research networks are systematically faster than the commercial networks. For the research network, performance is always more than three times higher for sites hosted in Europe and medRxiv. This is surprising for medRxiv, hosted by Cloudflare, as the other sites, including those hosted by Cloudflare, show much lower values and a very acceptable RTT.

The performance of commercial networks is higher and less stable throughout the month. The RTT for the same site varies greatly, up to several hundred milliseconds between the best and worst RTT recorded. In contrast, the RTT from the research network varies by only a few tens of milliseconds between the best and worst times.

From 23 January and for a week, latency peaks can be seen for most of the sites. On the other hand, for the research probe, from the 24th all the sites except the European ones followed the same trend. Initially, the RTT drops dramatically and then rises again a few days later.

Figure 4.2 shows the evolution of the RTT over the course of a day In IPv4. For each two-hour interval, the average RTT has been calculated from all the

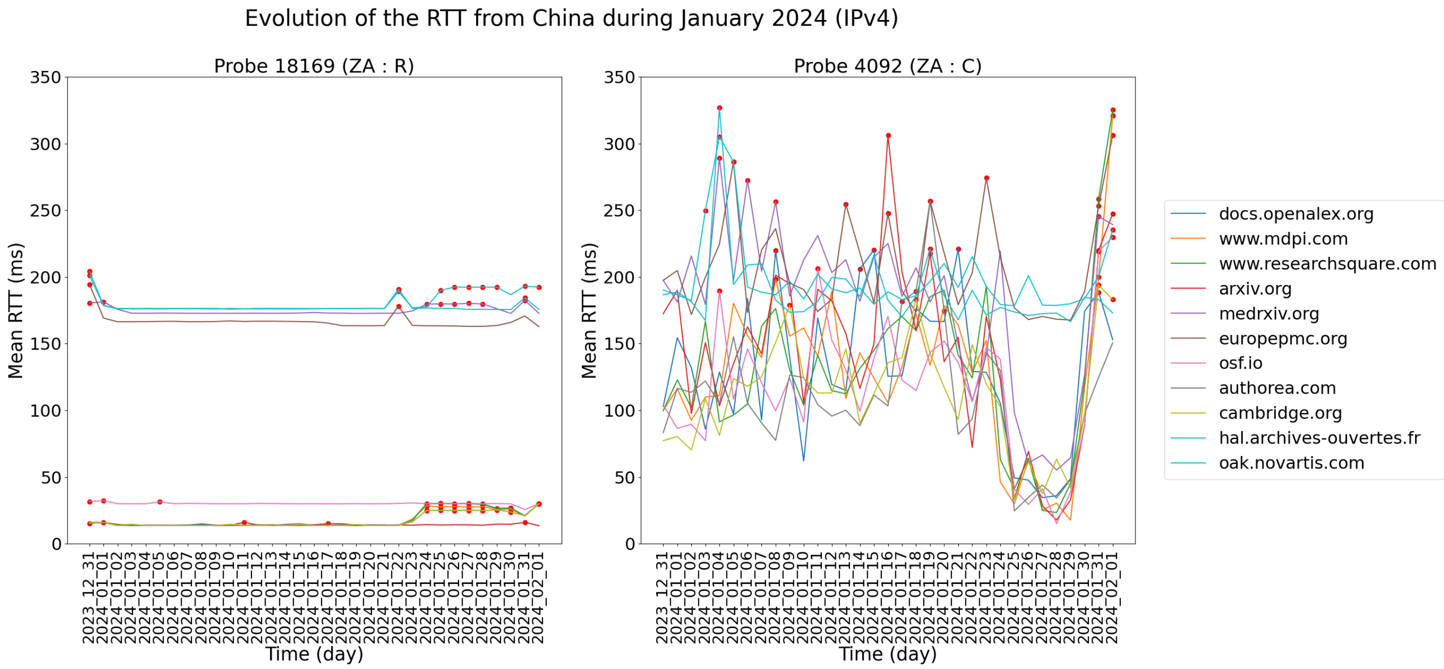

Figure 4.1: Comparison of the mean RTT from South Africa to repositories during January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

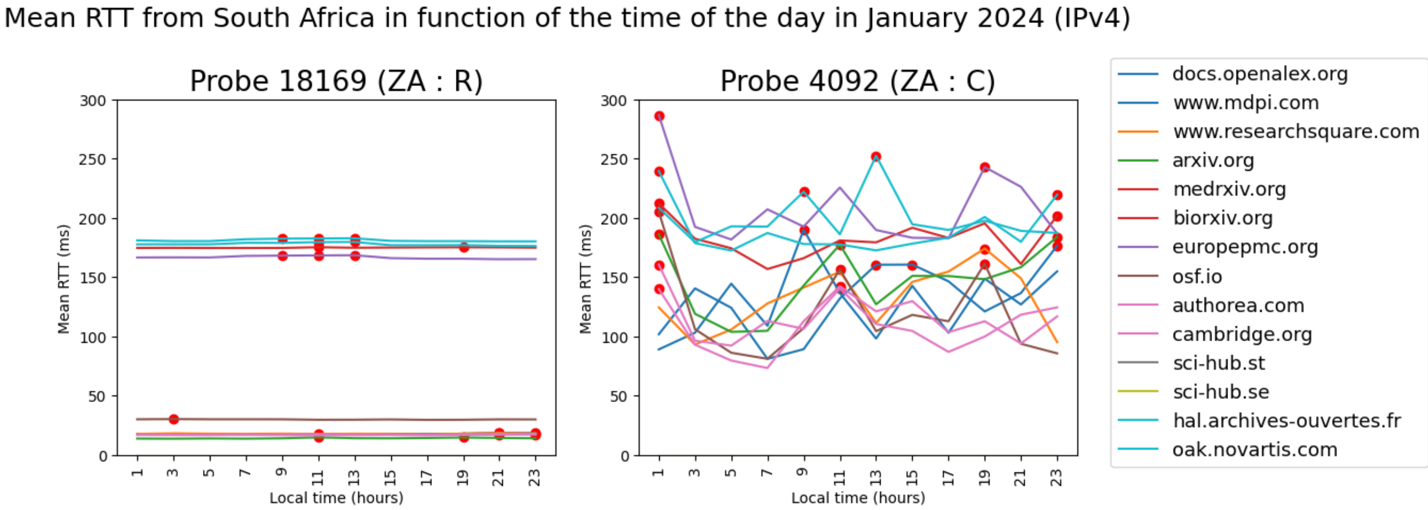

Figure 4.2: Comparison of the mean RTT from South Africa of open access repositories during in January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

measurements taken at that hour.

There is no significant peak for the research probe, the peaks detected are minimal and do not indicate a congestion problem due to heavy use.

For the commercial probe, the situation is different. There is much more variation over the course of a day, at most sites there are peaks between 11 pm and 1 am, and a few other peaks have been observed over the course of the day at some sites, although they are less obvious.

For IPv6, the performance of the research probe is similar for 3 sites, with very good performance in terms of RTT, but medRxiv shows much worse performance. In January 2024, the figure 4.4 shown the evolution of the mean RTT during Janviar 2024 from South Africa, a few peaks were observed, but nothing too alarming. However, the commercial probe saw an exceptional peak on 31 January at 1 am, when the average RTT reached up to 9,000 ms for 2 out of 4 sites. In the hours and days that followed, everything returned to normal, and the rest of the month remained stable.

During the day ash shown in figure 4.4, there is a peak around 9 am and several smaller peaks for several sites in the evening. For the research probe, a very large peak is recorded for 2 of the sites at 1 am, exceeding 2,000 ms, which is attributable to the peak on 31st of January. The rest of the day remained stable, with no significant peaks.

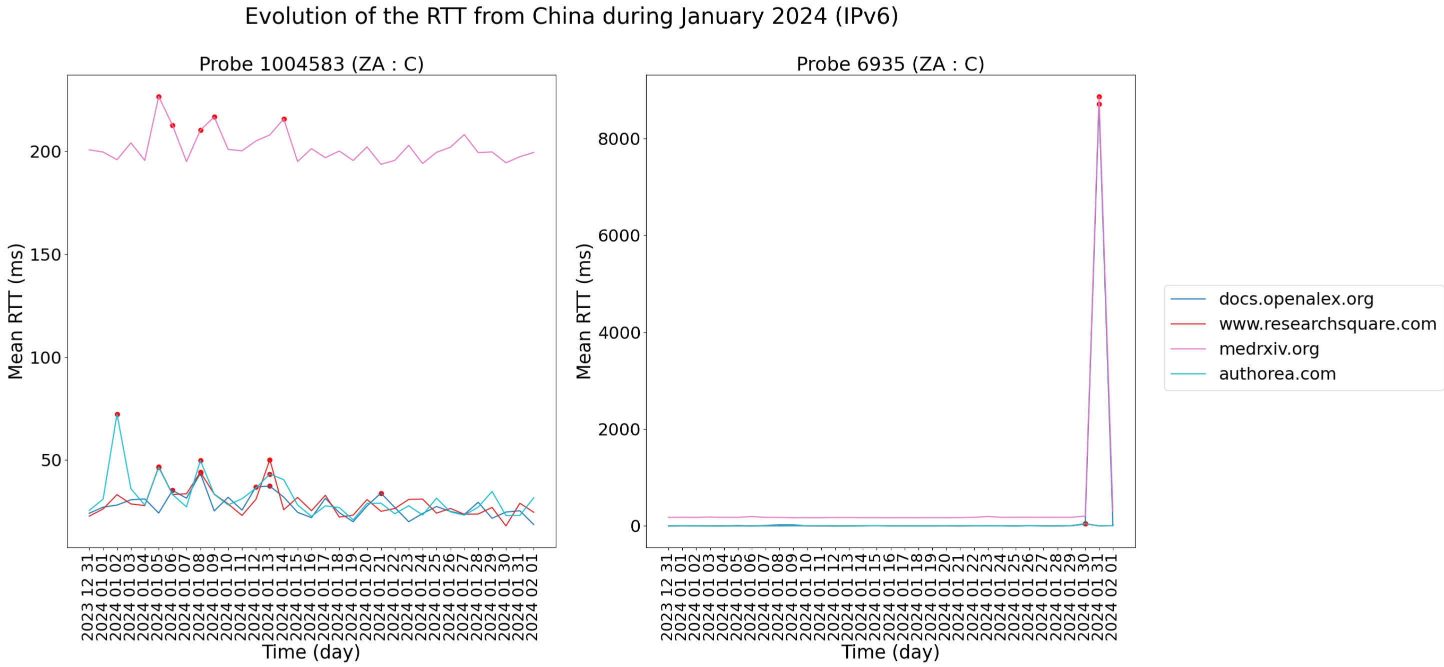

Figure 4.3: Comparison of the mean RTT from South Africa to repositories during January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

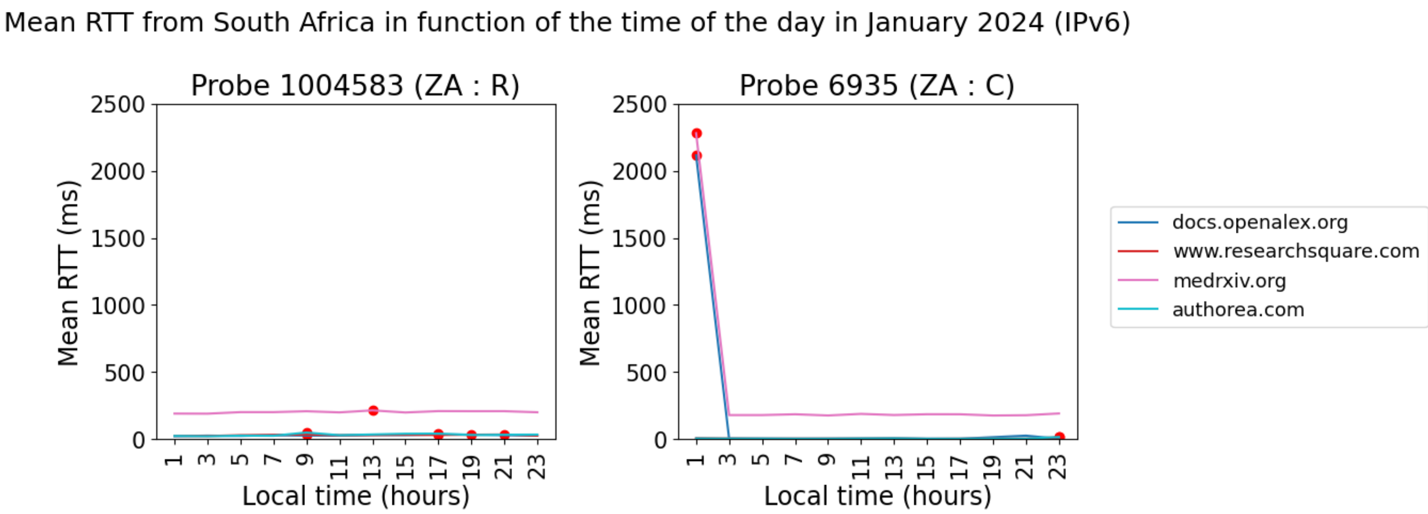

Figure 4.4: Comparison of the mean RTT from South Africa of open access repositories during in January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

The research probe generally travels fewer hops to reach its destination, but has a higher RTT. The number of hops remains stable and varies little for the same site. AS path also changed very little over the month, with only one or two changes observed for research and commercial probes.

The paths to reach the various CDNs are similar from the commercial probe. The packet passes through various providers in South Africa, then joins the host's network and arrives at the data centre. The path from the research probe passes through Tenet and quickly reaches the destination network.

The only site hosted by a CDN that does not follow this type of path is MedRxiv. For the commercial probe, two different routes were followed by the traceroutes. The first is via PCCW to the United Kingdom and join Cloudflare. The second route crosses the Seacom undersea cable via South Africa, Mozambique and Tanzania. This second route takes less time than the first, so it is to be preferred.

The traceroutes from the research probe to medRxiv also took two different routes. Unlike packets destined to other websites, the traffic does not always transit through Tenet. The first route observed is via Liquid Telecom to Marseille, then access Cloudflare. The second route crosses Tenet, but without much difference in RTT, although it is slightly lower. It seems that these routes are not stable, as over the course of the month several changes were observed between the two routes.

These routes explain the variation in RTT between MedRxiv and the other sites hosted by Cloudflare, which have a direct connection to the Johannesburg datacenter.

As we saw earlier, the RTT to the European sites is longer. When we look at the routes from the commercial probe, they all pass through London via PCCW before reaching their final destination.From the research probe, the path starts via Tenet and reaches GÉANT in the Netherlands, then GÉANT in the UK or France, depending on the destination country. From there, the packet crosses the research network of the local country. For packets destined for the Irish site, the traceroute passes through Tenet in South Africa, then joins Tenet in Amsterdam before heading for Dublin.

Another route has been observed, reducing the RTT by several tens of milliseconds. This route uses the TENET network to cross different countries: South Africa, the United Kingdom, the Netherlands (AMS), then arrives in Ireland. It uses the West Africa Cable System (WACS) which goes from Cape Town to London, whereas the previous route goes via Mtunzini, a town in South Africa. The Mtunzini-Amsterdam route takes much longer than the Cape Town-London route.

The routes taken by IPv6 vary very little and all pass exclusively through South Africa to reach Cloudflare's server in Johannesburg.

To summarise the case of South Africa, the country performs well compared to other countries in Africa and the world. This can be explained by the fact that most repositories are hosted in CDNs that have data centres in South Africa. For countries that do not have a server in Africa, such as sites hosted in Europe and medRxiv, performance is higher. The only way to reach these sites is via Europe or other African countries.

The research network probe consistently shows better IPv4 and IPv6 results than the commercial network probe. Apart from a few exceptional cases, there does not seem to be any congestion problem on the network.

4.2.2 China

China is the second most populous country and the second largest economy in the world. China has a very important Internet development policy and has reached 1.079 billion Internet users in 2023, which represents an Internet penetration of 76.4%, of which 70% live in cities and 30% in rural areas. The country has a very large IPv6 network and there are more than 767 million IPv6 users [23]. The Chinese government imposes censorship on many websites and despite a well-developed Internet infrastructure, it is not possible to access all the content desired. China is linked to the rest of Asia by several submarine cables, and to Europe, the United States and Australia by one cable. The country has 279 RIPE Atlas probes, but only around thirty are connected.

The Chinese probes selected both support IPv4 and IPv6, so these two probes took measurements for both protocols. This will make it easier to compare the results and rule out any differences due to hardware or probe location. The measurements, consisting of 12 pings and one traceroute per day, with a single packet per measurement, were carried out throughout January 2024.

The performance measured from the commercial and research networks to the open access directories in IPv4 are different as demonstrated in the figure 4.5. The commercial networks are more efficient in terms of response time, with an average of 2.8 ms over the month compared with 39 ms. However, the peaks observed are much higher from the commercial probe, rising to over 7,000 ms. The fact that these peaks only occurred once during the month suggests that this is an exceptional rather than a systematic problem. This is in contrast to the peaks in the measurements taken from the research probes, which are smaller but occur more often over the course of the month. These peaks can rise to several hundred milliseconds, which is not negligible when they are recurrent, as is the case with the research probe, so it seems to be more prone to congestion than the other probe. Figure 4.5 shows very clearly that the RTT to European sites: europepmc.org, hal.archives-ouvertes.fr and oak.novartis.com is systematically at least 3 times higher than those hosted in the United States.

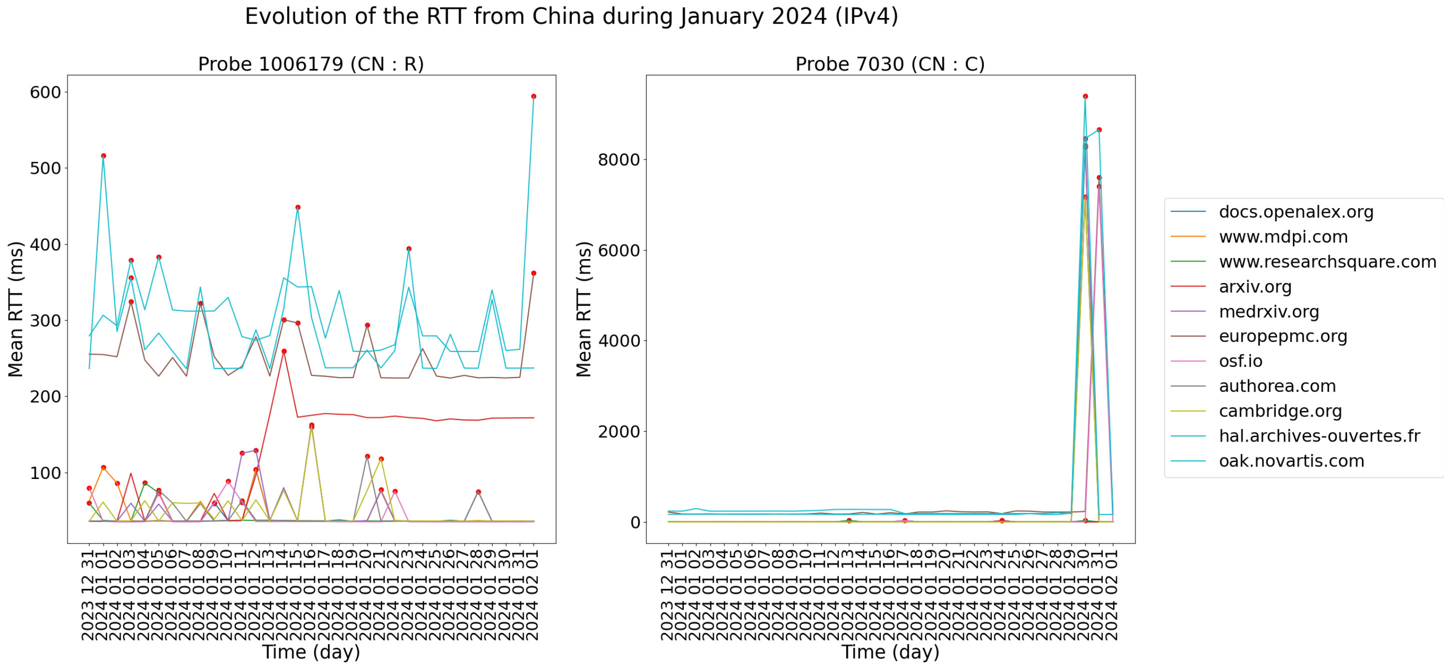

Figure 4.5: Comparison of the mean RTT of open access repositories during in January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

On a typical day in January as shown in figures 4.6, the commercial probe peaks appear at night or very early in the morning, usually between 11 pm and 7 am, depending on the site. In rare cases, there are peaks at 1 pm or 5 pm. These times correspond to non-work periods. For the search probe, the peak periods occurred throughout the day, although the most common period is between 11pm and 1am.

As with IPv4, there is a difference between response times measured from the commercial and research networks in IPv6. The average response time for the search probe is 42.4 ms for all the measurements made. For the commercial network probe, the average is 490 ms but, as shown in the figure 4.7, for the whole of January, the RTT is close to 0, except one or two exceptional peaks at the end of the month which, where the RTT rises to over 60,000 ms and 90,000 ms. Since this phenomenon was observed on all 4 repositories, it must be a technical problem or a heavy temporary congestion either on the probe side or the server side. If we exclude these unusual values, the average is 2.8 ms, which is an extremely efficient time and more representative of the entire month.

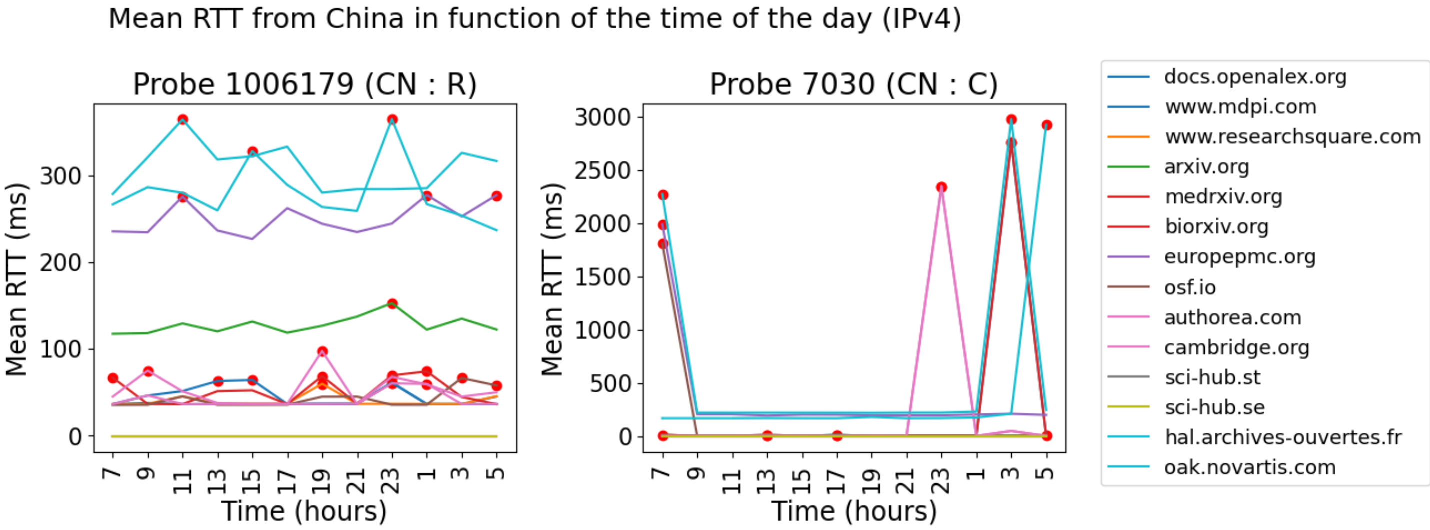

Figure 4.6: Comparison of the mean RTT of open access repositories from China during a typical day in January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

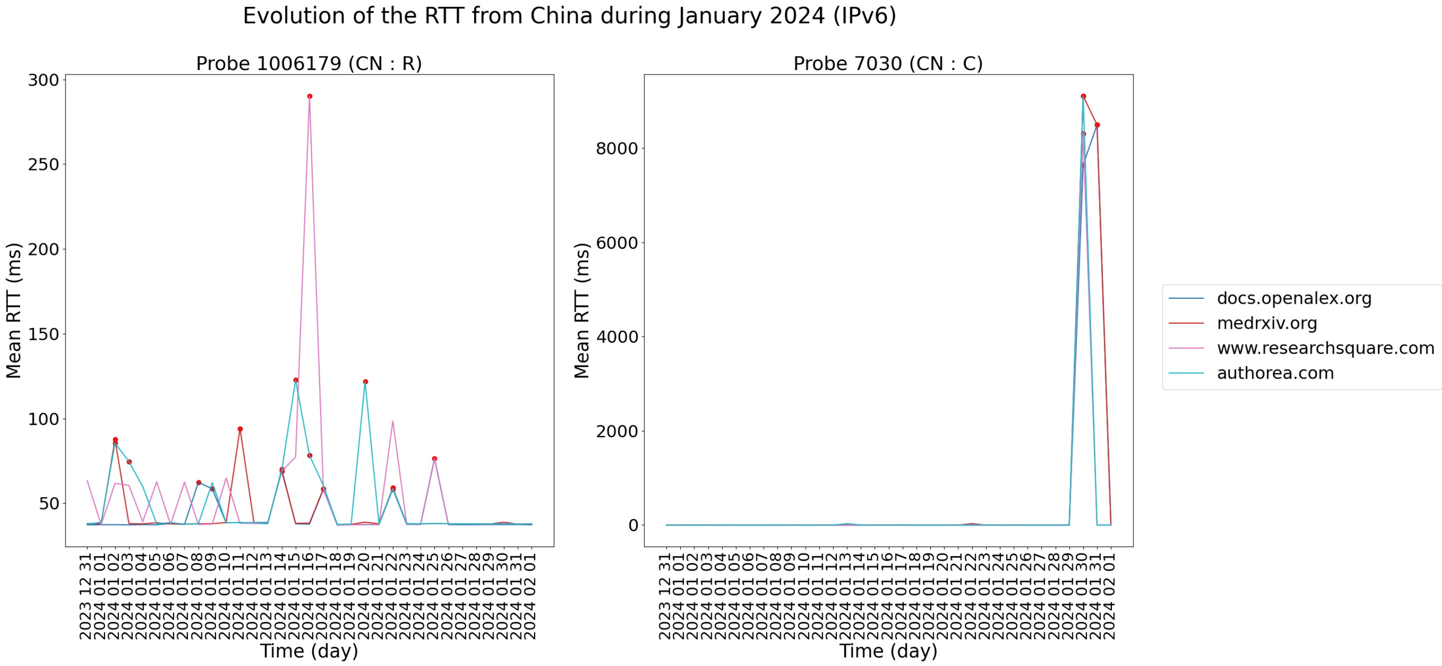

Figure 4.7: Comparison of the mean RTT of open access repositories during in January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

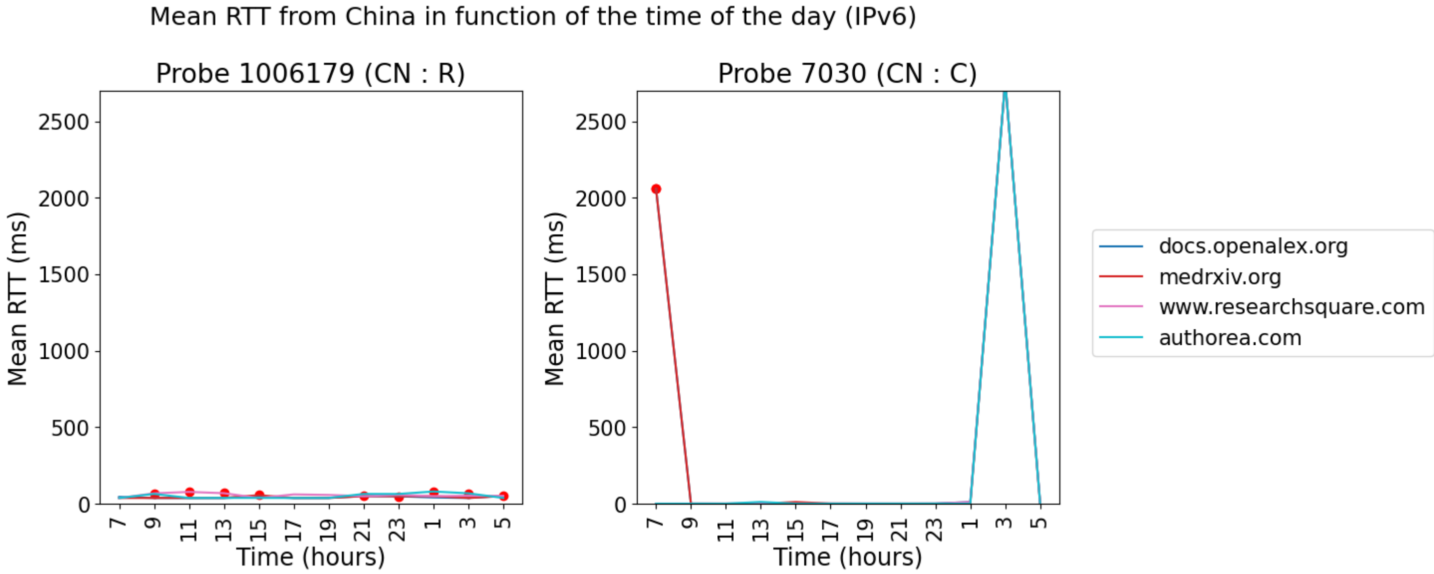

Figure 4.8: Comparison of the mean RTT of open access repositories from China during a typical day in January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

Figure 4.8 shows in detail the average response time as a function of local time on a typical January day in China for IPv6. The data represents the average RTT for each hour over all the days in January, excluding values of over 60,000 ms. This gives a more accurate idea of the variations in RTT over a typical day of the month.

With regard to the commercial probe, depending on the site, two peak periods can be identified, with very little variation in values for the rest of the time. The first is during the night, at 1 am and 5 am. The second is in the early afternoon, at around 1 or 3 pm. However, even during the peak, the RTT at these times is around 12 ms, which is still very low.

There is more variation in the measurements made by the commercial probe. Some peaks are identifiable, but at different times depending on the site. Three periods of peaks can be distinguished: in the morning around 9 o'clock, which corresponds to the arrival of employees at the office, in the early afternoon around 3 o'clock and during the night between 9 pm and 3 am.

Over the course of the month, when comparing the two IP versions, some peaks may occur on the same day for the search probe, but this is not always the case. For example, when the destinations are Researchsquare or Openalex, the peaks occur at different times. On the other hand, for the commercial probes, the high peaks on 30 January are observed in both IPv4 and IPv6, which suggests a problem affecting the infrastructures of the network.

The peaks during the day do not occur at the same time between IPv4 and IPv6 from the same probes and to the same destinations.

The RTT to European sites are more than twice as high as to the sites host by a CDN, which can be explained by the path taken by the packets. When we study the AS paths, the packets destined for Europe, several paths are possible.

The first is the one used by packets going to the United Kingdom and France, which follow a similar pattern. From the commercial probe for England, the packet passes through Hong Kong and arrives in London, the destination city, via the Russian-European Transit Network (RETN). For France, the packet passes through Hong Kong and arrives in France via RETN as well, the packet goes into IXP managed by Renater, the French research network, which also hosts the hal.archives-ouvertes site. For the research probe, the path goes through CERNET, the Chinese research network, to cross China. To arrive in the UK in London, one of the two European destinations of Orientplus. Orientplus is a European and Chinese project to link the research networks of the two regions. The packet then joins either Janet, the UK research network if the destination is Europepmc, or the GEANT network in London. Once in the GEANT network, it arrives in Paris and passes through the Renater network to the destination.

A second route was observed on 29 January for both destinations. First, the packet goes from CERNET to the CNIC-C, Computer Network Information Centre of Chinese Academy of Sciences. It then goes to GEANT in Marseille (France), then Geneva in Switzerland and returns to Paris. If the packet is destined for Europepmc, it joins Janet, and if it is destined for hal.archives-ouvertes, it joins Renater. On 30 January, the situation returned to normal with the first route. This second route is therefore an alternative route when there is a problem with Orientplus.

The second case is towards Ireland. The AS path also starts in Hong Kong, but instead of going to Europe, it heads for the United States and then to Dublin for both types of probes. From 17 January and for the rest of the month, there was a change of route for the commercial probe, resulting in a higher hop number but a lower RTT. The route then passes through London and Paris to arrive in Ireland, and no longer through the United States.

Thanks to traceroutes and the results of DNS lookups, it has been deduced that the Chinese probes accessing directories hosted by CDNs are connected to servers located in Hong Kong. For these sites, the paths are similar whether for a commercial or search probe. The packet goes to Hong Kong either in CERNET for search probes or in a commercial network where it is handled by different networks depending on the destination: Cloudflare, Fastly, Google, Arelion, Tata telecommunications, etc.

Regarding IPv6, the paths taken are similar to IPv4 for the four Cloudflare sites. With the exception that for the search networks, the packet passes through CERNET2, CERNET's IPv6 network and then joins the Cloudflare network in Hong Kong, it then follows the same path as for IPv4.

In summary, IPv4 performance from the commercial probe is more efficient than from the research probe. The research probe is more prone to congestion over the course of the month. On a typical day, the research probe shows peaks in RTT throughout the day, varying from site to site. The commercial probe, on the other hand, shows clear peaks in the evening and at night. IPv6 performance is weaker, with significant RTT peaks for the commercial probe. RTT peaks are observed for all IPv6 sites from the commercial probe during the night. The RTT to European sites is much higher than that the ones hosted by CDNs with servers in Hong Kong.

4.2.3 India

India is the most populous country in the world, with 881.255 million internet users in March 2023, representing 62.6% of the total population, mainly in rural areas [50]. The vast majority of internet use is on mobile phones, thanks in particular to the development of 5G [9]. The country is linked by numerous submarine cables to the rest of Asia, Europe and Africa, and has one of the largest terrestrial fibre networks in the world. India has nearly 300 RIPE Atlas probes installed, with more than half connected (May 2024).

The measurements made from India were carried out throughout the month of January 2024, the pings were made every 2 hours with one packet per measurement, the traceroutes were made at the same time, with the same number of packets once a day at midnight in Brussels. The measurements were made from 3 different probes, the research probe made the measurements for IPv4 and IPv6, two different commercial probes were used to make the measurements for IPv4 and IPv6.

As shown in figure 4.9, for IPv4 research networks, the results for CDN sites are very good, with an RTT of around ten milliseconds, while European sites show a significantly higher RTT than the others. Over the course of the month, a few peaks were observed for measurements taken from the research probe; on the 22nd, half the sites experienced a peak, although these were not considerably higher. Apart from these occasional peaks, the RTT is relatively stable.

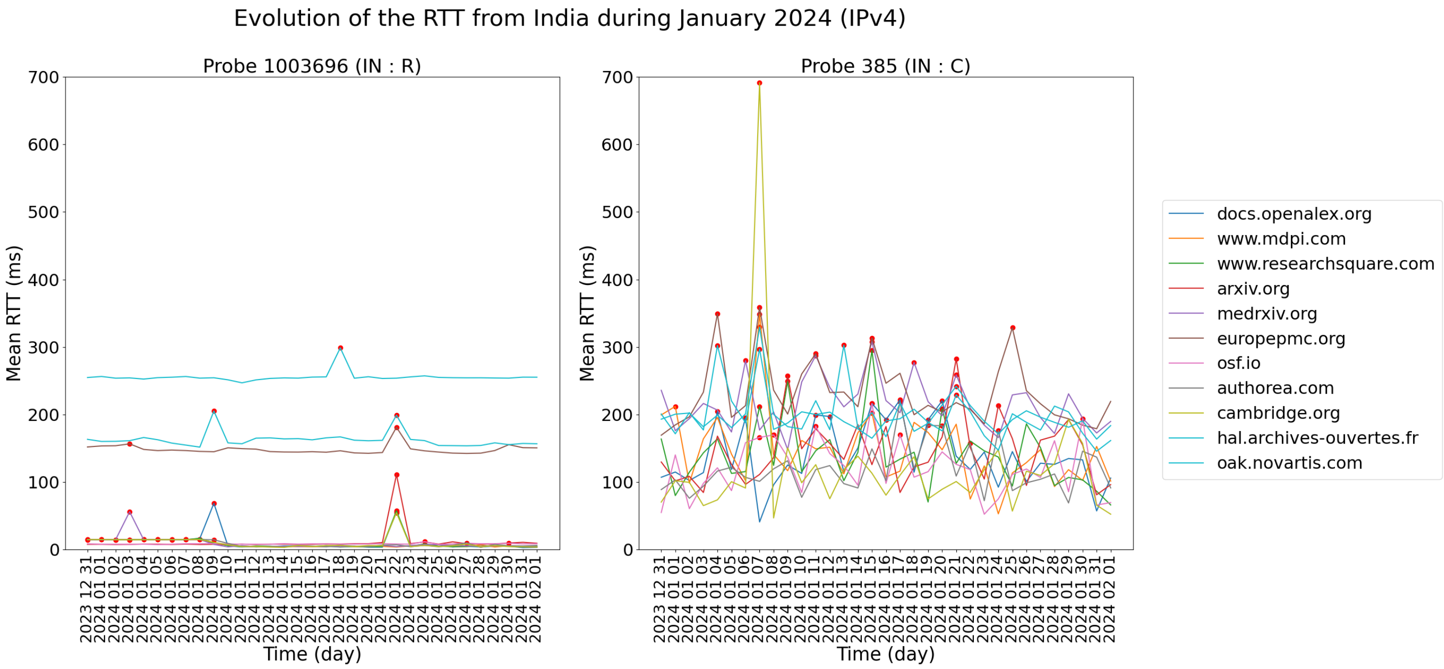

Figure 4.9: Comparison of the mean RTT from Inde to repositories during January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

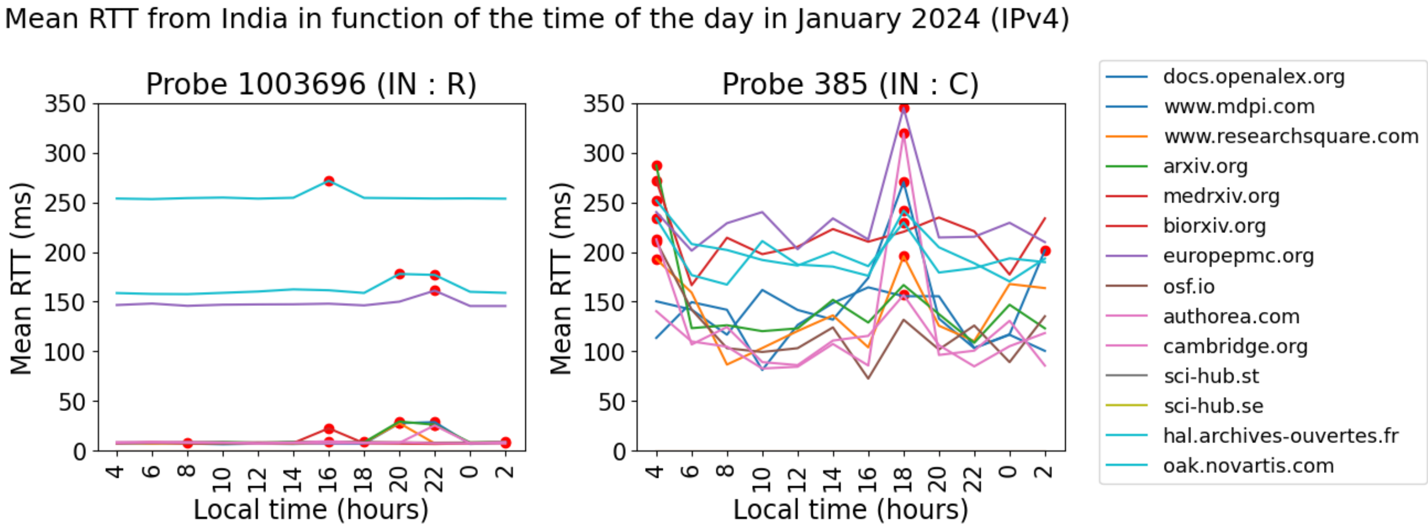

Figure 4.10: Comparison of the mean RTT from Inde to repositories during a typical day in January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

On the other hand, for the commercial probe, the RTT is higher than for the research probe and varies considerably for all sites throughout the month.

Figure 4.10 shows a typical day in January 2024, obtained by averaging the measurements collected over the month at a given time.

For the research probe, the RTT is very stable throughout the day, with a slight peak in the late afternoon and early evening for some sites, but these peaks are not very high. Unlike the commercial probes, which observe peaks for all sites at 6 pm and 4 am. This is probably due to congestion caused by many people using the network at the same time. The RTT also varies much more over the course of a day.

Figure 4.11 shows the evolution of the average RTT over the month for the four directories supporting IPv6. IPv6 performance is similar between the research and commercial probes, except for medrxiv, which shows a much higher RTT on the commercial probe. The commercial probe showed a higher RTT until 10 January, when performance improved significantly for the rest of the month for all directories, except a notable peak for one site on the 22nd.

For the commercial probe, there are three high peaks for Research Square, particularly towards the end of the month.

On a typical day as shown in figure 4.12, the peaks do not occur at the same time for IPv4 and IPv6. In IPv6, there is no time when all four sites show peaks

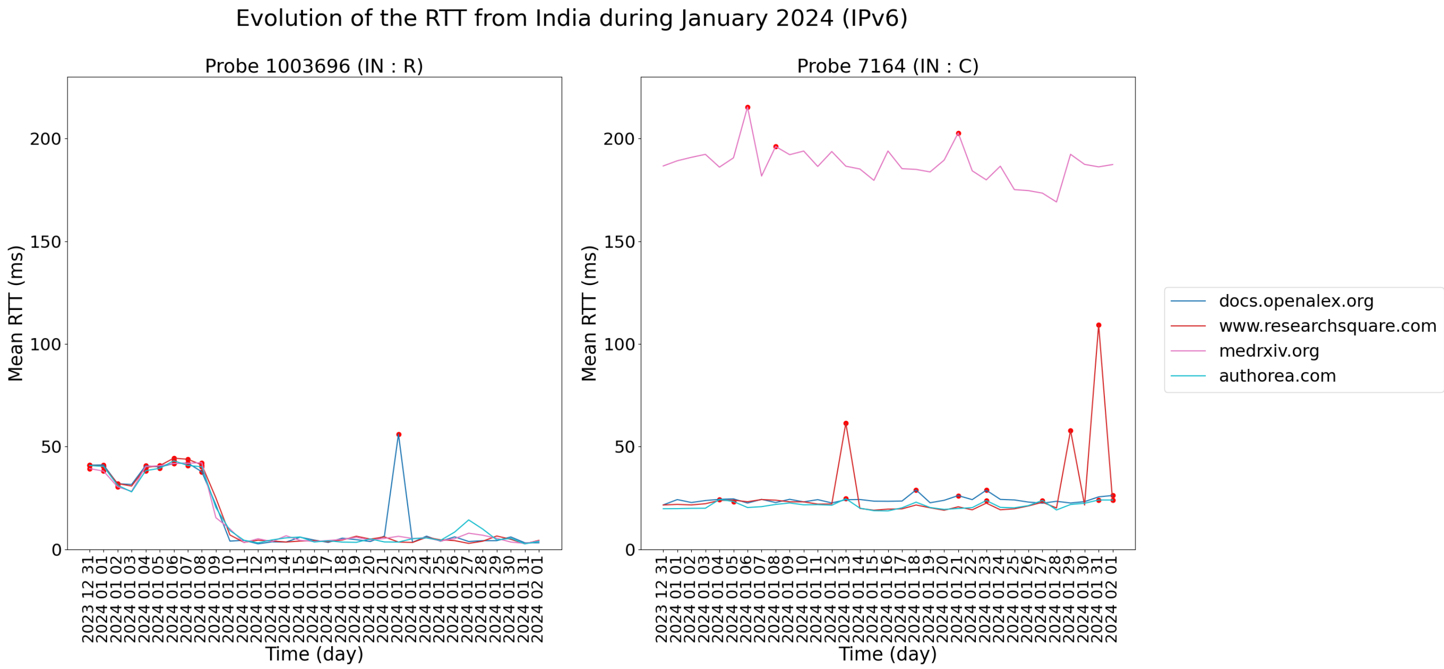

Figure 4.11: Comparison of the mean RTT from Inde to repositories during January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

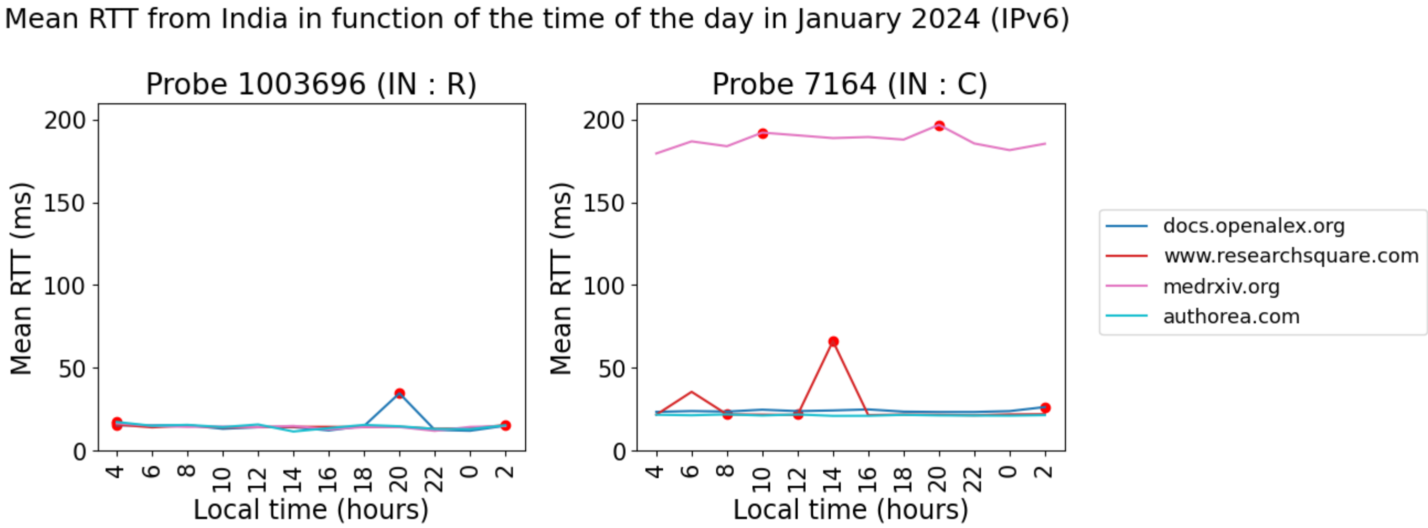

Figure 4.12: Comparison of the mean RTT from Inde to repositories during a typical day in January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

simultaneously. For the research probe, there is a high peak at 8 pm for one site, while for the commercial probe, a peak is observed at 2 pm for another site.

For IPv6 measurements coming from the commercial probe, the AS path goes through Airtel, the performance is very good and the number of hops is low. As for the research probe, the traceroutes gave very little information about the path followed. However, on 10 or 11 January, depending on the destination sites, we saw a major drop in RTT and the number of hops due to a change in the AS path. The AS path still only passes through India, but via a different supplier. After this change, the AS path remains the same for the rest of the month.

In the case of MedRxiv, the RTT is significantly higher than for the other sites hosted by the same CDN, reaching more than a hundred milliseconds in the best case and more than 800 in the worst case compared with around ten milliseconds for the other sites. The number of jumps is also slightly higher. The route passes through Frankfurt in Germany.

We can therefore assume that the servers used to host MedRxiv are located in Frankfurt or in Europe, which would explain the high performance to this destination and the results similar to those of the European sites. This result is surprising, given that the IP address advertised is the same as that of Cloudflare's other sites, which are hosted on servers in India.

The AS paths used to reach France from the commercial probes and the research probes both pass through the RENATER network. They arrive via Marseille and then return to Paris.

To reach London, the commercial probe packets also pass through Marseille, this time via Twelve99. They then pass through Paris, Frankfurt, Hamburg and finally London, where they join JANET before reaching their destination. The research probe follows the same route, except that it goes directly from Paris to London, which saves a few milliseconds compared with the route taken by the commercial probe. The names of the cities were determined using the codes in the DNS response and the Twelve99 website [85], which indicates which city each code corresponds to.

To reach Ireland and OSF.io, the commercial probe packets pass through Airtel and arrive in Amsterdam, then enter the Blacknight network. For the research probe, it is more difficult to determine the exact location of the path taken because of the lack of information on the IPs in the traceroute.

In IPv6, packets follow the same AS paths as those in IPv4, including MedRxiv, which also passes through Frankfurt.

To summarise, in IPv4, research networks offer better performance than the commercial probe. The sites hosted in Europe and MedRxiv have a higher RTT than the others, because they are hosted in Europe, unlike the other sites which are hosted in India and are therefore closer. Over the month, a number of disparate peaks were identifiable for the research network, while for the commercial probe, the trend was much less clear with peaks all the time depending on the site. Congestion phenomena were observed at two times of the day for all sites for commercial probe IPv4 at 4 am and 6 pm. For the IPv4 research probe, there were a few peaks during the day, especially in the afternoon and evening, but these varied depending on the site.

In IPv6, performance between the research network and the commercial network is very similar, with the exception of one site which performs less well on the commercial network. For the research probe, the RTT is much higher at the beginning of the month, then there is a change of routes and a decrease in RTT for all sites. The commercial probe shows some peaks during the month. On a typical day, there are peaks at the beginning of the day, to all 4 sites for the research probe, while for the commercial probe, few peaks occur at different times of the day for some sites.

4.2.4 New Zealand

As one of the islands of Oceania, New Zealand is geographically distant from the other continents, which has a major impact on inter-continental internet traffic. The island is linked by a few submarine cables to the USA and other countries in Oceania.

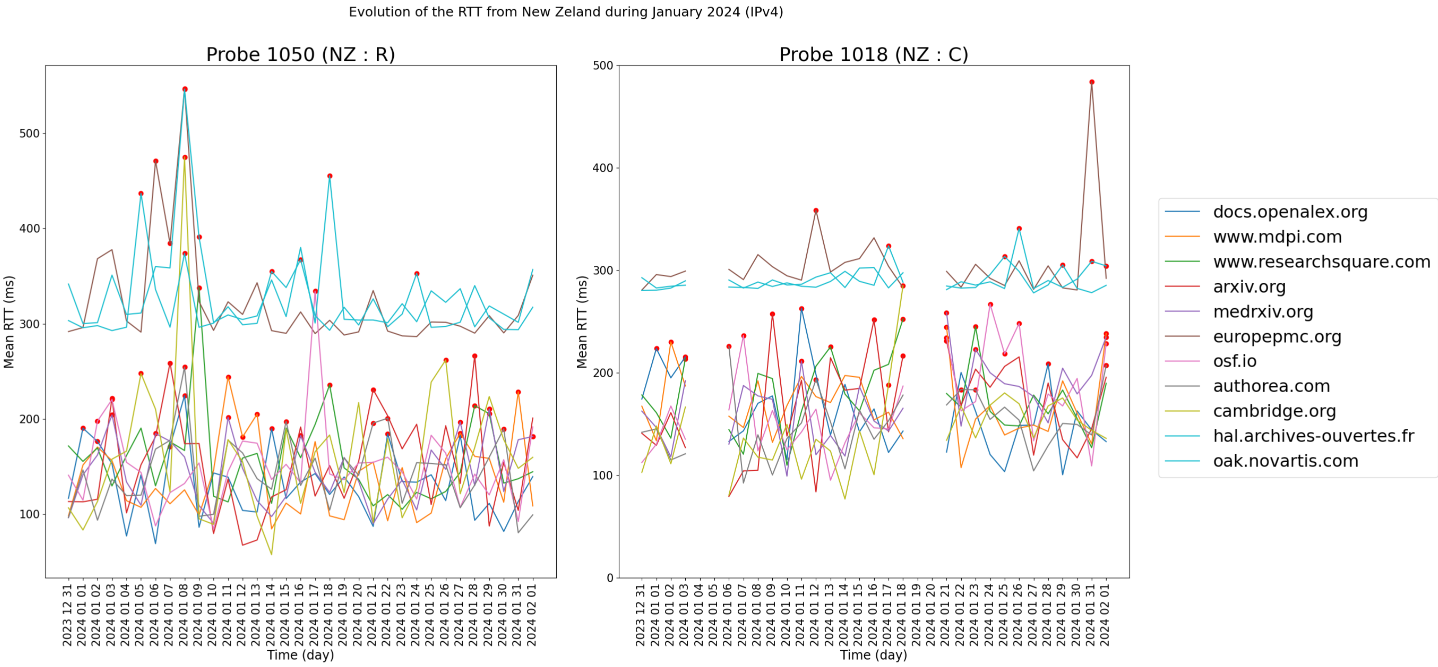

Figure 4.13: Comparison of the mean RTT from New Zeland to repositories during January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

Concerning the internet within the country, nearly 5 million inhabitants are internet users, which means that the penetration rate is 95.9%, one of the highest in the world [80, 54]. New Zealand has almost a hundred probes installed and 3/4 of them were connected on 4 May 2024.

Performance from New Zealand was measured throughout January 2024. Pings were performed at 2-hour intervals, and traceroutes were performed once a day at midnight. The measurements were made using the same probes for IPv4 and IPv6. It should be noted that the commercial probe did not provide any results for 4 days out of the month.

In IPv4, as we can see in figure 4.13 performance is very high overall, with an RTT of around 100 ms in the best case, and over 500 ms in the worst case. The performance of the research and commercial probes is similar in terms of magnitude and varies enormously over the month, with numerous peaks and no clear trend. The three European sites once again show a higher RTT than the others, with an RTT of around 300 ms in the best case. For the 4 repositories which deployed IPv6, the performances are similaire to the one in IPv4 as shown in figure 4.14.

Over the course of a day, variations are also considerable as shown in figure

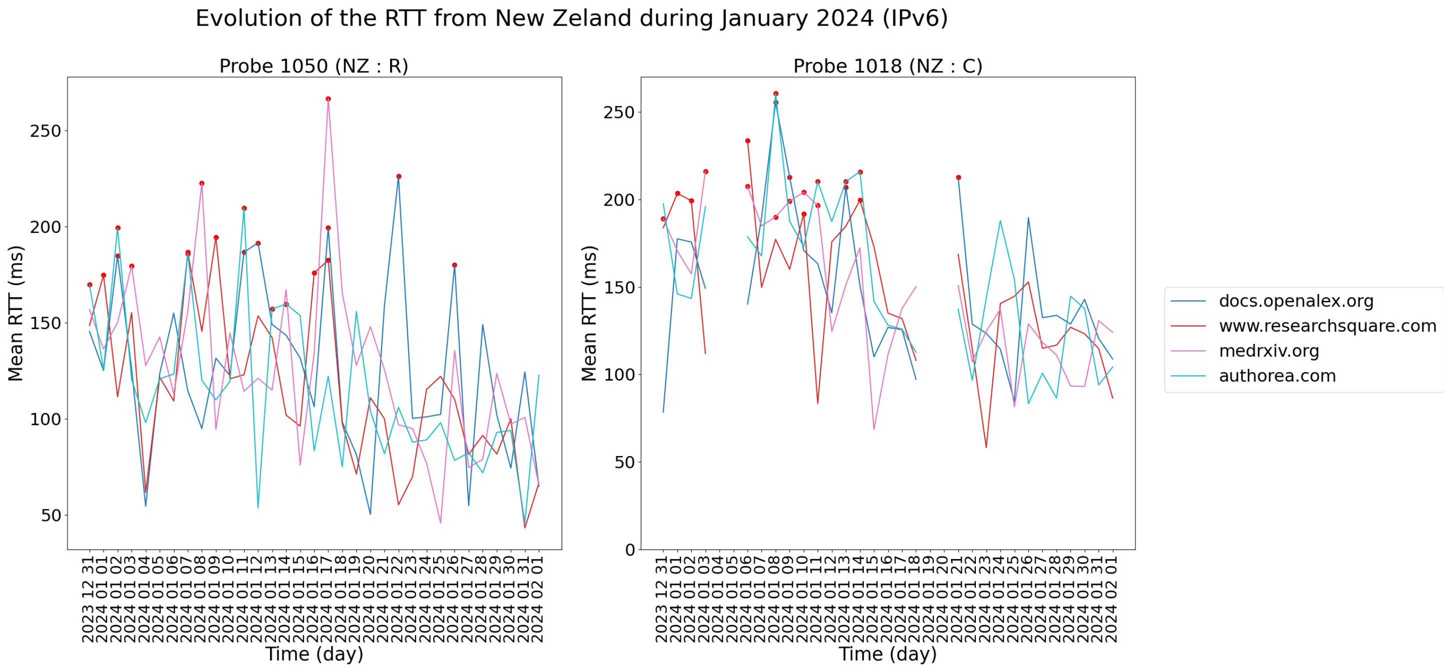

Figure 4.14: Comparison of the mean RTT from New Zeland to repositories during January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

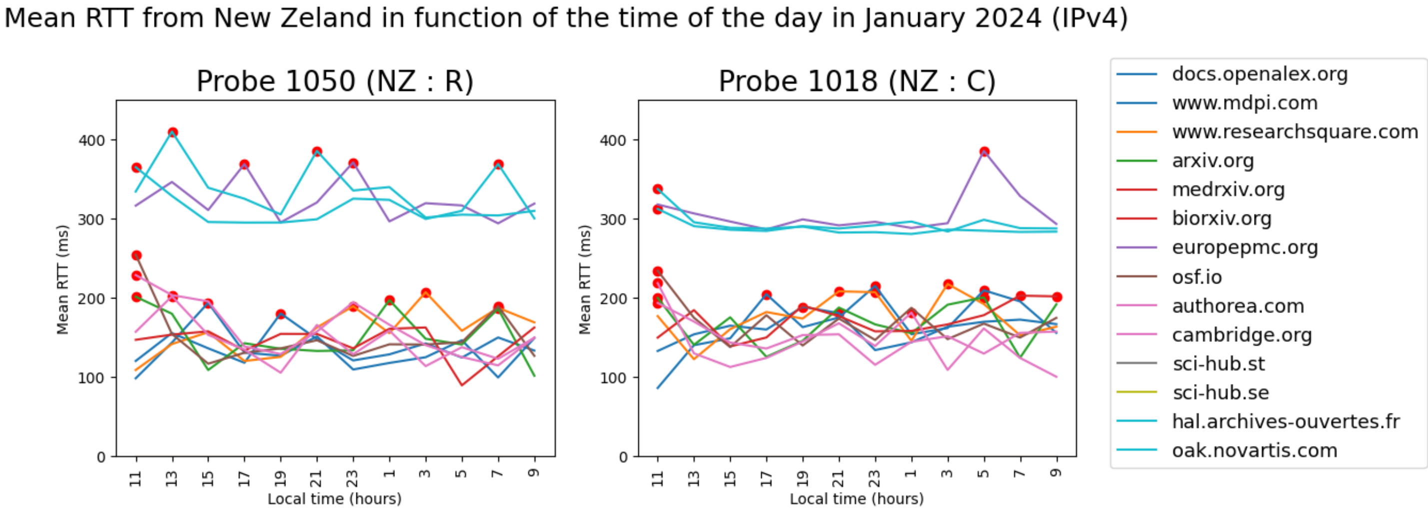

Figure 4.15: Comparison of the mean RTT from New Zeland to repositories during a typical day in January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

4.15 : sites show peaks at different times, making it difficult to conclude that there is any particular congestion. This is also true for IPv6 as shown by figure 4.16. At certain times of the day, we see both the best results and the worst results for different sites.

Content hosted by a CDN accessed from New Zealand is hosted in New Zealand or Australia. This conclusion is drawn from the traceroutes carried out and the DNS responses associated with the traceroute IP addresses. We were able to observe that most of the ASes present in the AS paths originating from a commercial probe in IPv4 belonged to the same company Voyager.nz. Thanks to the DNS lookup responses in which the code of the names of the cities passed through is indicated, it was possible to deduce that the AS path passes through several New Zealand cities before entering the host's network; although the cities sometimes change, the AS path remains the same. For some sites, the AS path begins in the same way, but then heads towards Australia.

From the IPv4 research probe, the AS path passes through REANNZ, the New Zealand research network. Similarly to the commercial probe, it is also possible for packets to be redirected to Australia instead of staying in New Zealand.

The traceroutes to Europe PMC provide limited information on the IP addresses traversed by the packets. We can see the first hops in New Zealand, either at Voyager for the commercial probe, or at REANNZ for the research probe, as well as the final destination. Between these two points, there is no information available

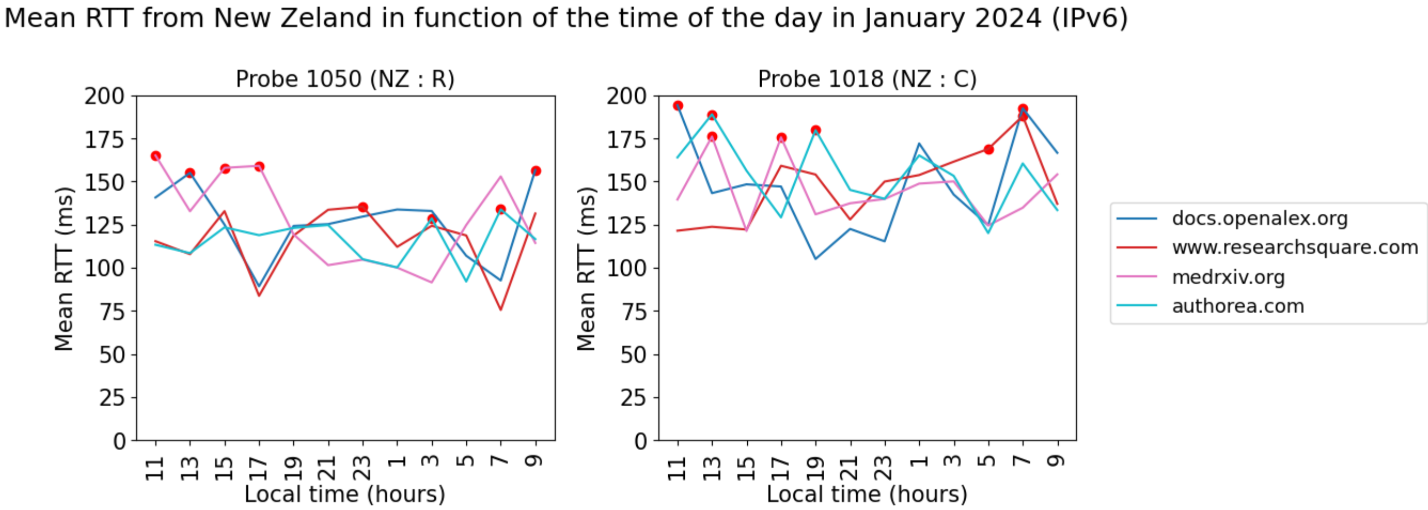

Figure 4.16: Comparison of the mean RTT from New Zeland to repositories during a typical day in January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

for either the research probe or the commercial probe.

For packets destined for France, the commercial probe uses Voyager to cross New Zealand, then arrives in the United States via GTT in Palo Alto, before reaching Paris and finally entering the RENATER network. For the research probe, there is no information available between REANNZ and the final destination.

Traffic to Oak from the commercial probe also passes through Palo Alto in the USA with Twelve99 after crossing New Zealand with Voyager. It then heads for New York before reaching Dublin and finally Blacknight.

As for the research probe, the traffic passes through REANNZ: North Shore, Mangawhai, Honolulu, Seattle, then there's no more information until Blacknight. The route taken in the REABBZ network has been confirmed using the information available on their site [72] in addition to the city codes present in the DNS lookup response.

In conclusion, performance is high overall, particularly for sites not hosted by CDNs, which is explained by the fact that the routes used to reach Europe must first pass through the United States. In terms of latency, the research and commercial probes have similar results for IPv4 and IPv6, but the variations in latency do not occur simultaneously. We have observed that IPv6 and IPv4 latency measurements also perform similarly for the four sites that support both protocols. In general, the situation is rather disorderly: all the sites behave differently, but remain within the same order of magnitude in terms of latency.

4.2.5 Argentina

The last country analysed is a South American country called Argentina. Over 89.2% of the population are netizens in 2024, representing 41 million people [49]. The country is linked to the other countries on the American continent by just 3 submarine cables. The country has around fifty probes deployed, just over half of which are connected.

The pings from the probes were made throughout January 2024, 12 times a day every 2 hours, while the traceroutes were made daily at midnight GMT+1. Each measurement was made by sending a packet. The measurements were taken from 3 probes, the research probe made the measurements in IPv4 and IPv6, while the other two are part of a commercial network, one made the IPv4 measurements and the other the IPv6 measurements.

Performance is better from the research probe, with an average of 9 ms over all measurements, compared with 122 ms for the commercial probe. However, as for the other countries, performance to sites hosted in Europe is worse, varying between 200 and 400 ms. This trend has been observed with both probes. The medrxiv site has similar performance to those hosted in Europe, even though it is hosted by Cloudflare, as shown in figure 4.17 which illustrates the evolution of the RTT during the month of January 2024 for the commercial and research probes in IPv4.

From the research probe, the RTT is low and regular over the month, with the exception of a few peaks, particularly at the end of the month. On the other hand, from the commercial probe, the RTT is higher and varies considerably, with numerous peaks spread throughout the month at all sites.

Figure 4.18 shows the RTT over the course of a typical day in IPv4 where peaks are observed at different times of the day for each site, both for the research and commercial probe, suggesting that there is no pattern of periodic congestion.

For IPv6, the average total RTT for all measurements is identical between measurements made using the commercial and research probes, but the distribution over time is different. Figure 4.19 shows the evolution of the RTT during January 2024 for the two IPv6 probes. The research probe has a RTT lower than 13 ms throughout the month except for one day when the average for the day is very high. On the other hand, the commercial probe has a RTT that is always around 13 and 14 ms in the worst case. In both cases, this indicates very stable performance.

Figure 4.20 shows a typical day in January in IPv6, there is a peak at 2 pm for the research probe, corresponding to the peak on January 30 towards Research Square. The other sites show slight peaks between 2 and 4 am, as well as 10 am and 2 pm. The research probe shows a very high degree of stability, with variations in RTT of less than a millisecond depending on the time of day, indicating normal

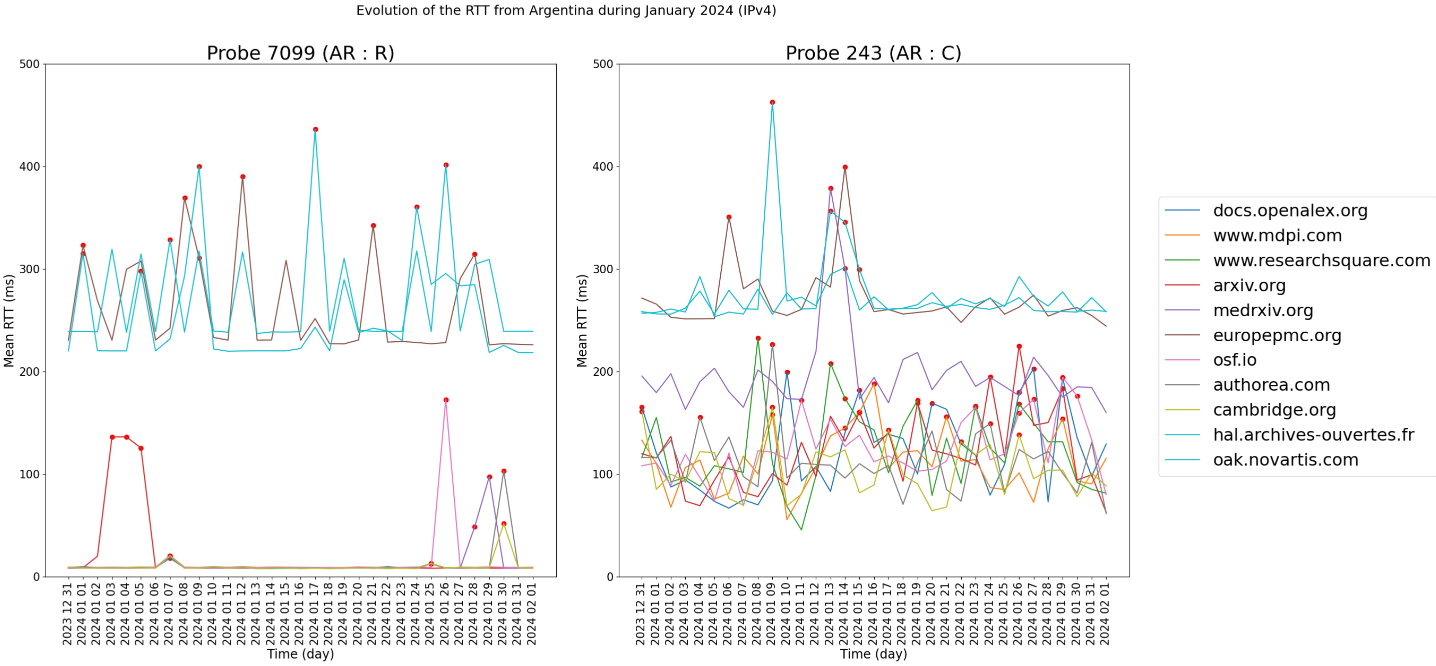

Figure 4.17: Comparison of the mean RTT from Argentina to repositories during January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

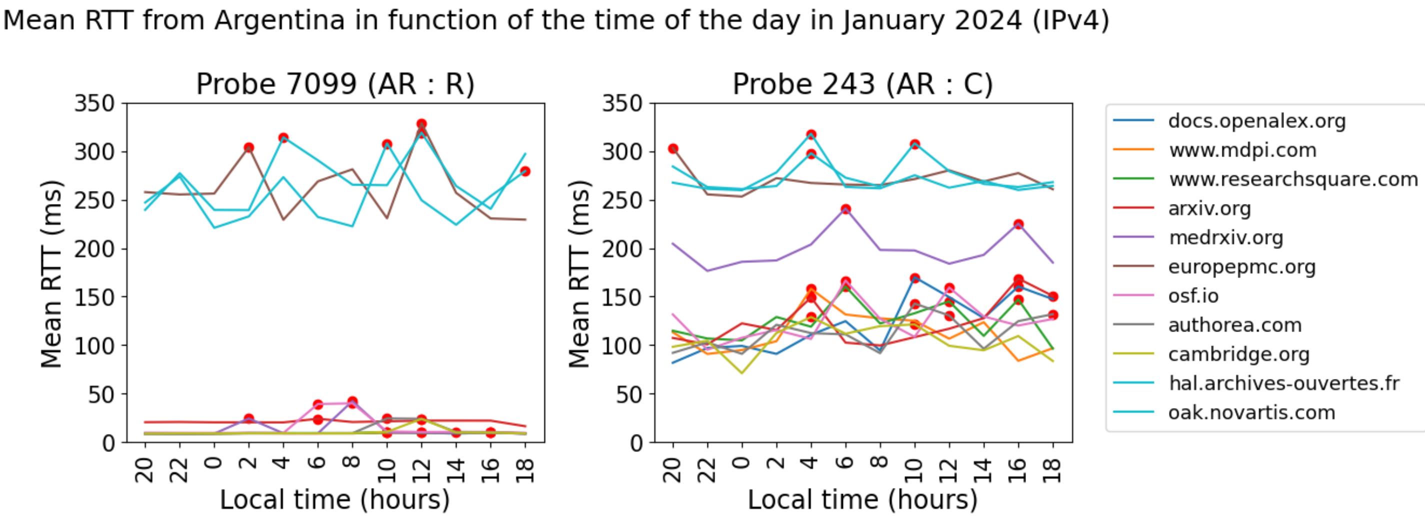

Figure 4.18: Comparison of the mean RTT from Argentina to repositories during a typical day in January 2024 in IPv4. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

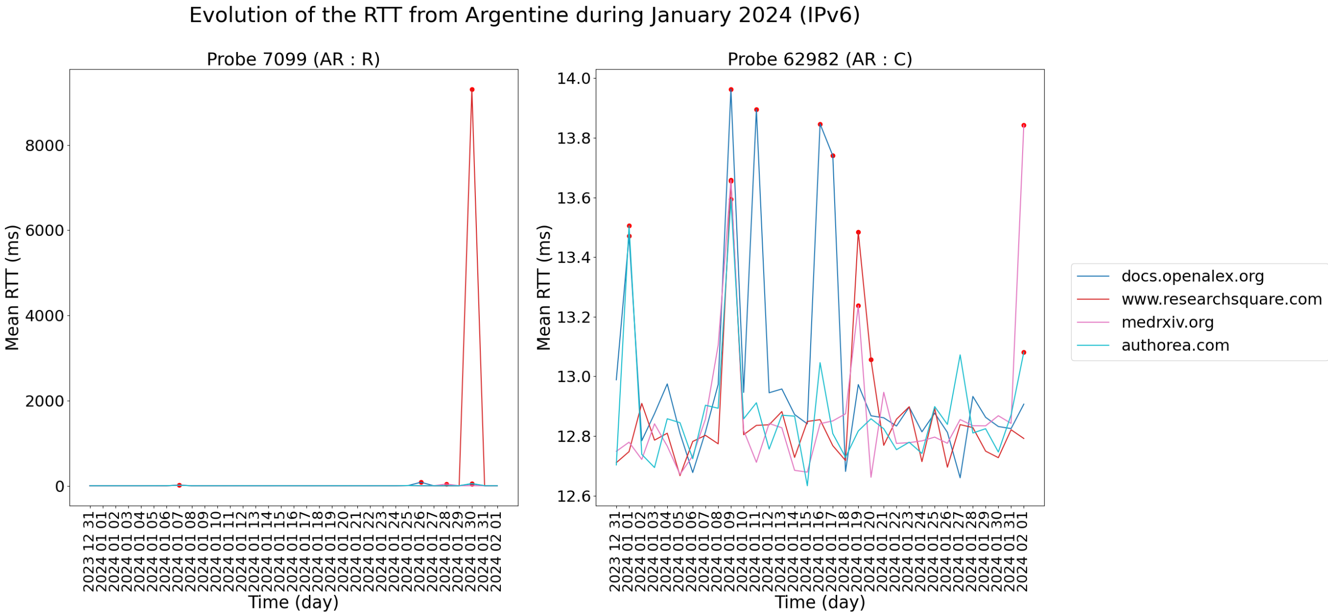

Figure 4.19: Comparison of the mean RTT from Argentina to repositories during January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

fluctuations in traffic rather than episodes of congestion.

The AS paths obtained using IPv4 traceroute in January 2024 are stable over the month for all sites. They reveal that packets from commercial probes to open access repositories hosted by CDNs go to servers in Argentina, except MedRxiv where packets transit via the United States to Miami, which explains the increase in RTT compared with the other sites.

All intra-Argentina traffic from the commercial probe is routed via Telecom Argentina, the country's largest operator. On the other hand, from the research probe, the packets pass through the network of a major Argentina research centre, Centro Regional de Investigaciones Básicas y Aplicadas (CRIBABB), to reach data centres in Argentina.

To reach the sites hosted in France and the United Kingdom, the commercial probe goes via Telecom Argentina before reaching the United States to arrive in Europe, i.e. Paris, London or Dublin, depending on the destination.

The research probe passes through CRIBABB and joins the RedCLARA network in Chile, as CRIBABB is not part of RedCLARA. Then, the route goes through the link between RedCLARA and GÉANT to arrive in Portugal and crosses several European cities: Madrid, Marseille, Geneva, Paris. Once in Paris, the route joins either JANET or RENATER, depending on the destination. To reach Ireland, the

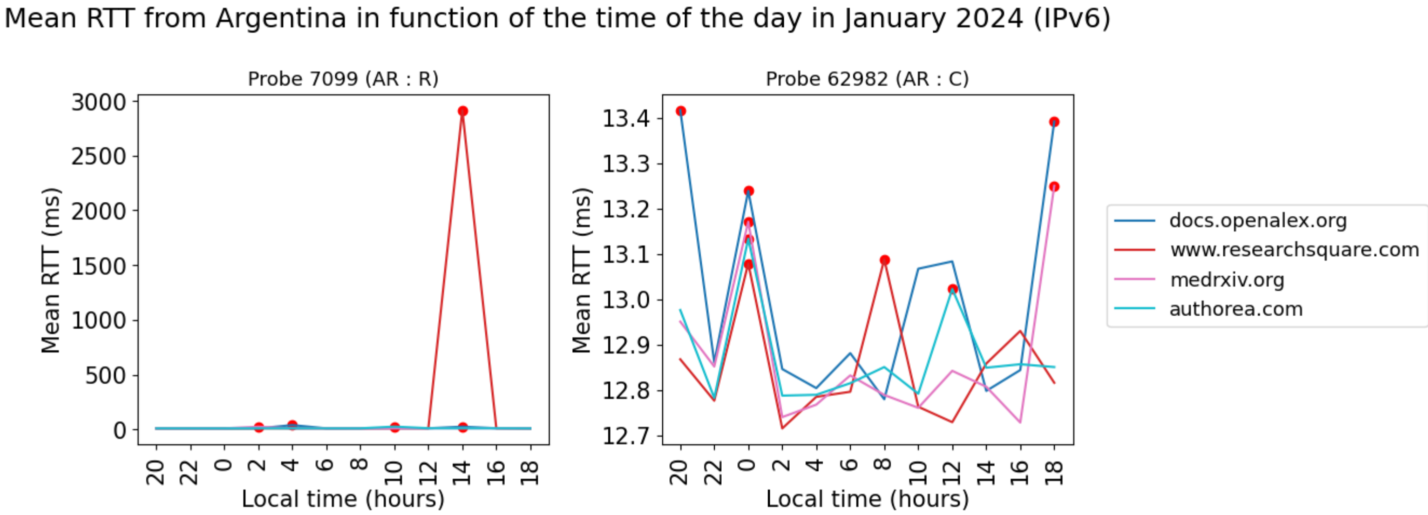

Figure 4.20: Comparison of the mean RTT from New Zeland to repositories during a typical day in January 2024 in IPv6. The red points shown on the graph represent peaks in the RTT, calculated from the standard deviation with a degree of freedom of 2.

route is not the same, the packets pass through CRIBABB but go through the United States via the Lumen Technologies network to arrive in Dublin. It probably does not pass by RedCLARA and GÉANT because the site is not hosted on a research network, contrary to HAL and Europe PMC.

In IPv6, measurements taken in the direction of the 4 sites hosted by Cloudflare using the commercial probe show the same AS path. The same phenomenon is observed using the research probe. The number of hops remains constant over the month and is the same for all sites.

In conclusion, performance from Argentina is generally low and shows little variation for directories hosted by CDNs. In IPv4, the research network is more efficient than the commercial network, with the exception of European sites where performance is equivalent. With IPv6, performance is almost identical between the two probes.

Since most of the sites are hosted on national servers, the AS path passes mainly through Argentina. For the commercial probe, it passes through Telecom Argentina, while for the research probe, it passes through CRIBBAB. However, for sites hosted in Europe and MedRxiv, packets from the commercial probe pass through the United States, while those from the research probe pass through the RedCLARA and GÉANT organizations.

Chapter 5

Recommendations

During our analysis, we identified a number of areas where improvements could be made for the accessibility of open access content on a global scale.

As we have seen, one of the most decisive factors for performance is the hosting of the site by a CDN. We therefore strongly recommend to sites that don't use a CDN to integrate one, as the differences in performance between those that do and those that don't are significant.

We are advising CDNs to install data centres in regions of the world where they have little or no presence. We have observed major disparities between regions that have such infrastructures and those that do not, resulting in longer and more circuitous transmission routes, particularly in Africa. Corneo's article [27] has the same conclusion, recommending the installation of more data centres and servers in regions that need them.

In terms of global and intercontinental connectivity, it is crucial to strengthen the submarine cable infrastructure between continents, particularly from Latin America and Africa. However, it is essential to put in place cautious routing policies so as not to compromise other efficient routes that are not affected, as has been seen in the past [38].

We advise to all governments to create policies to promote the development of Internet infrastructure as it was the case in Latin America with good results. By legislating, investing, and collaborating with private sectors, research institutes and education [16].

Local governments have the power to create policies to attract private investors to build infrastructure and create new services closer to areas that lack them, as Chetty recommends [22].

For Africa, it is recommended to strengthen intercontinental connectivity as well as land connections between African countries that are underdeveloped and complicate inter-African trade. This would reduce dependence on intercontinental routes and improve intra-continental performance. It is also necessary to install more data centres in the region to improve performance and to install more RIPE Atlas probes to better study performance within African countries and to better identify areas that need improvement.

The performances we observed for the Asian continent are generally good for most countries. However, in India and China we have seen daily congestion peaks. It is therefore advisable to study these peaks and implement measures such as load balancing to avoid these recurring congestion problems. In addition, it has been noted that less efficient routes are sometimes used for a period of time, as it happened in India for the research probe. We therefore recommend that optimal routes be used to improve accessibility to open access content.

It is essential to use the most efficient routes between continents. For example, it is much quicker to use the Cape Town-London route than the Mtunzini-Amsterdam route, both of which are currently used from South Africa. This implies a small deviation via London to get to the Netherlands, but since these two cities are well connected, the time saving is still significant.

In Europe, performance to open access content is generally the best, due to the very advanced development of its network infrastructure and the presence of numerous data centres. However, measurements taken from other parts of the world, particularly from commercial networks, reveal very high RTTs. Although research networks are performing slightly better, thanks to GEANT's many partnerships with other research networks around the world, latencies are still high. However, this ensures more stable routes and lower latency than commercial networks. We therefore recommend developing more similar partnerships. In addition, we advise European sites to use CDNs so that copies of their sites are quickly available anywhere in the world, not just in Europe.

North America, and in particular the United States, hosts a multitude of powerful Internet infrastructures, including numerous data centres, which translates into good performance for the country. We recommend continuing to develop connectivity with other continents, as the USA is often the centre of Internet routes and frequently serves as a transit point between continents.

For South America, we recommend that local governments continue the Internet infrastructure development policies that are currently in place [16]. These efforts must be supported and amplified to maximise their impact. It is also crucial to strengthen partnerships with other continents and improve intercontinental connectivity.

It is particularly important to increase the number of submarine cables between

Latin America and Europe. Currently, there is only one direct cable dedicated to communication between GÉANT and RedCLARA, forcing packets from Latin America to transit through the United States to reach Europe. This has been observed in traceroutes from Argentina to EuropePMC, OAK and HAL. Increasing direct connectivity between these regions would reduce latency and significantly improve overall performance.

Chapter 6

Conclusion

This master thesis investigated the Internet accessibility of open access content across the world's research and commercial networks in IPv4 and IPv6. The aim is to identify connectivity issues and performance inequalities. It is important to be able to access this content, particularly in developing countries, to facilitate the development of new knowledge.

The first step was to identify the repositories hosting open access content, as well as their hosts and locations. This involves identifying whether the host is a CDN and if it utilizes anycast. The second step was to select the most stable RIPE Atlas probes for different countries.

One of the limitations of our study resides in the use of RIPE Atlas, which is subject to a significant bias in the distribution of probes around the world. Many probes are deployed in Europe and the United States, but relatively few in developing regions. These regions have less developed Internet infrastructures and generally inferior performance, but it is difficult to study them in depth due to the lack of probes to carry out measurements there.

Then, we carried out measurements throughout January 2024. Pings were made every two hours, and traceroutes once a day. Each measurement consisted of sending a packet from probes located in research and commercial networks around the world.

We then looked at RTT on a global scale, comparing performance between the different continents. First without distinguishing between research and commercial networks, and then by separating them. First for IPv4, then for IPv6.

Next, we deepened the analysis in five specific countries: South Africa, China, India, Argentina and New Zealand. First, we studied the evolution of RTT over the course of a month, and then over the course of a typical day. Finally, we examined the AS paths and routes taken by the packets.

The measurements revealed inequalities between continents, with connectivity more effective in Europe, while Africa and Oceania performed less satisfactorily. Disparities were also observed within continents, such as South Africa, which achieved good results, unlike the rest of Africa, which is often forced to go via other continents to reach sites hosted by CDNs.

We also found disparities in performance between probes located in research networks and those located in commercial networks within the same country. On some continents, the commercial networks perform better, while on other continents it is the research networks that perform better.

We observed daily congestion peaks for China and India, while for the other countries analysed, no specific pattern was identified. Large peaks were also observed over the course of the month. For some countries, such as the commercial network in China, these peaks were exceptional. By contrast, other countries, such as New Zealand and Argentina, showed much more recurrent peaks, suggesting a more serious congestion problem.

We have observed that performance is highly dependent on the host of the open access directory. Sites hosted by CDNs using anycast with servers spread around the world, generally perform well. On the other hand, sites that are not hosted by CDNs, in particular the sites hosted in Europe in our study, perform much less well.

The reason for this is that traffic has to travel a greater distance, in regions where connectivity is not highly developed. In addition, intra-continental connections can also be slow in some regions, because infrastructures are not as developed as in Europe and the United States. Research networks have well-developed links with each other, including between continents, but the RTT remains high.

We have therefore made a number of recommendations based on our observations. As we have seen, the main factor influencing performance is hosting by a CDN. We therefore recommend that sites not yet hosted by a CDN should do so. In addition, we suggest that CDNs install servers in regions of the world that lack them, particularly in Africa, where there are very few.

To support this initiative, local governments could put in place policies to attract internet investors and businesses. In addition, we encourage governments to adopt funding policies and legislative measures, as it has been done successfully in Latin America shown by Carisimo [16], to improve internet infrastructure and therefore overall performance.

It is also crucial to improve connectivity between continents and between countries within continents, particularly in Africa, both intercontinental and intra-continental. Particular attention needs to be paid to the choice of routes taken, as there are often more efficient routes that are not used, or are only used late, as it has been observed in India.

We recommend to network managers to analyse daily congestion patterns and plan measures to mitigate them. For example, in China, highly identifiable congestion peaks occur at the same times every day. One solution could be to implement load balancing to effectively distribute the traffic load and improve network performance.

Studying the Internet accessibility of open access content is important, but the key is to make the results of research available in Novartis OAK. As shown by the European University Association (EUA) in 2021, 85% of scientific publications were published in paid-for journals [33], making them inaccessible to many people, particularly in developing countries that do not have the budget to pay these fees. Therefore, as well as improving connectivity to open access repositories, it is above all essential to promote open access culture.

Chapter 7

Appendices

7.1 Use of AI tools and translator

To improve the quality of the English version, ChatGPT was used to help correct errors in certain English sentences and to check turns of phrase. Deepl was used to help translate the text in English.

7.2 Github repositories

All the programmes carried out, the data collected and the results obtained in the context of this master thesis are available at the following link : `https://github.com/manonjac/memoire`

7.3 Tables and figures

Continent	country	number probe	pourcentage	Connected probe
Total Africa		381	2.24	219
	DZ	3	0.02	1
	AO	9	0.05	8
	BJ	12	0.07	5
	BW	5	0.03	2
	BF	12	0.07	8
	BI	3	0.02	1
	CM	5	0.03	2
	TD	2	0.01	1
	CG	0	0	0

	CD	7	0.04	3
	CI	1	0.01	1
	CF	0	0	0
	DJ	1	0.01	1
	EG	6	0.03	3
	EH	0	0	0
	ER	0	0	0
	SZ	1	0.01	1
	ET	2	0.01	1
	GA	1	0.01	1
	GM	0	0	0
	GN	0	0	0
	GH	6	0.03	5
	GW	0	0	0
	KE	22	0.13	16
	LS	1	0.01	1
	LR	0	0	0
	LY	0	0	0
	MG	7	0.04	5
	MW	3	0.02	2
	ML	5	0.03	2
	MR	0	0	0
	MA	8	0.05	4
	MZ	4	0.02	3
	NA	4	0.02	4
	NE	0	0	0
	NG	9	0.05	2
	RW	5	0.03	2
	RE	49	0.3	38
	SN	4	0.02	2
	SC	2	0.01	1
	SL	2	0.01	1
	SO	0	0	0
	ZA	135	0.78	104
	SS	1	0.01	1
	SD	3	0.02	0
	TZ	5	0.03	5
	TG	3	0.02	2
	TN	11	0.06	5
	UG	10	0.06	5

	ZM	9	0.05	6
	ZW	3	0.02	2
Total Antartica		1	0.01	0
Antartica	AQ	1	0.01	0
Total Asia		2859	16.21/16.45	1830
Asia	AE	25	0.14	20
	AF	7	0.04	7
	AM	18	0.1	13
	AZ	11	0.06	7
	BH	15	0.09	8
	BD	8	0.05	5
	BT	1	0.01	1
	BN	3	0.02	3
	KH	3	0.02	2
	CN	279	1.61	38
	GE	18	0.1	17
	ID	141	0.81	99
	IN	299	1.73	163
	IR	143	0.83	86
	IQ	11	0.06	7
	IL	51	0.29	39
	JP	347	2	259
	JO	3	0.02	3
	KZ	68	0.39	56
	KW	4	0.02	3
	KG	10	0.06	9
	LA	2	0.01	1
	LB	8	0.05	4
	MY	36	0.21	26
	MN	14	0.08	8
	MV	6	0.03	6
	MM	3	0.02	2
	NP	11	0.06	8
	OM	3	0.02	2
	PK	11	0.06	4
	PS	4	0.02	1
	PH	88	0.51	59
	QA	4	0.02	1
	RU	734	4.24	539
	SA	25	0.14	18

	SG	180	1	115
	KR	43	0.25	27
	LK	3	0.02	2
	SY	0	0	0
	TW	44	0.24	44
	TJ	10	0.06	6
	TH	40	0.23	29
	TL	1	0.01	1
	TR	93	0.54	60
	TM	0	0	0
	UZ	21	0.12	15
	VN	9	0.05	7
	YE	1	0.01	0
Total Europe		10062	58.37	8153
Europe	AL	32	0.18	20
	AD	5	0.09	5
	AT	294	1.70	251
	BY	31	0.18	23
	BE	237	1.37	192
	BA	24	0.14	14
	BG	131	0.76	106
	HR	45	0.26	35
	CY	18	0.1	15
	CZ	400	2.31	322
	DK	147	0.85	126
	EE	32	0.18	29
	FI	168	0.97	126
	FR	1346	7.77	1054
	DE	2234	12.89	1829
	GR	122	0.7	93
	GG	1	0.01	1
	GQ	2	0.01	1
	HU	71	0.41	57
	IS	16	0.09	13
	IE	125	0.72	97
	IM	5	0.09	3
	IT	542	3.13	379
	LV	47	0.27	40
	LI	5	0.09	5
	LT	45	0.26	42

	LU	61	0.35	46
	MT	11	0.06	10
	MD	21	0.12	18
	MC	1	0.01	0
	ME	8	0.05	7
	NL	807	4.66	635
	MK	15	0.09	10
	NO	126	0.73	97
	PL	253	1.46	183
	PT	317	1.83	217
	RO	102	0.59	82
	SM	1	0.01	1 0
	RS	68	0.39	54
	SK	58	0.33	45
	SI	62	0.36	56
	ES	290	1.67	222
	SJ	1	0.01 1	
	SE	267	1.54	214
	CH	430	2.48	363
	UA	247	1.43	182
	GB	809	4.67	622
	AX	1	0.01	1
Total North America		2979	17.18	2215
	BB	1	0.01	0
	BS	0	0	0
	BZ	1	0.01	1
	CA	394	2.27	304
	KY	4	0.02	3
	CR	6	0.03	4
	CU	1	0.01	1
	CW	1	0.01	1
	DO	18	0.10	14
	SV	2	0.01	0
	GL	2	0.01	1
	GP	3	0.02	3
North America	GT	4	0.02	1
	HT	3	0.02	2
	HN	9	0.05	6
	JM	1	0.01	0
	MQ	2	0.01	1

	MX	68	0.39	38
	NI	16	0.09	9
	PA	44	0.25	31
	PR	2	0.01	2
	BL	2	0.01	1
	KN	1	0.01	1
	LC	3	0.02	1
	SX	1	0.01	0
	TT	2	0.01	2
	US	2386	13.77	1786
	VI	2	0.01	2
Total Oceania		475	2.74	348
Oceania	AU	335	1.93	247
	CK	2	0.01	0
	FJ	4	0.02	2
	PF	6	0.03	5
	GU	7	0.04	4
	MU	13	0.08	6
	FM	2	0.01	2
	NC	6	0.03	5
	NZ	97	0.56	75
	PG	1	0.01	1
	TO	1	0.01	0
	VU	2	0.01	1
	SB	0	0	0
Total South America		416	2.42	283
South America	AR	53	0.31	29
	BO	27	0.16	11
	BR	136	0.78	93
	CL	72	0.42	59
	CO	23	0.13	15
	EC	20	0.12	11
	GF	1	0.01	1
	GY	0	0	0
	PY	9	0.05	7
	PE	12	0.07	9
	SR	1	0.01	1
	UY	20	0.12	17
	VE	42	0.24	30
Total		17102	100	12855

Table 7.1: Number of RIPE Atlas probes for each continent on 4 May 2024. The number of probes includes connected and deconneted probes (excluding abandoned probes).

Repositories	Host at	IP	Location
Open Alex	Cloudflare	2606:4700:4400::ac40:93d1 2606:4700:4400::6812:282f 172.64.147.209 104.18.40.47	Anycasted Anycasted Anycasted Anycasted
MDPI Preprints	Cloudflare	104.18.25.151 104.18.24.151	Anycasted Anycasted
Research Square	Cloudflare	2606:4700:3108::ac42:2921 2606:4700:3108::ac42:2adf 172.66.41.33 172.66.42.223	Anycasted Anycasted Anycasted Anycasted
arXiv	Cornell University	128.84.21.199	New York, US
(updated)	Fastly	151.101.195.42 151.101.67.42 151.101.131.42 151.101.3.42	Anycasted Anycasted Anycasted Anycasted
medRxiv	Cloudflare	2606:4700:3031::ac43:a6ec 2606:4700:3033::6815:bb4 104.21.11.180 172.67.166.236	Anycasted Anycasted Anycasted Anycasted
bioRxiv	Cold Spring Harbor Laboratory	143.48.220.223	New York, US
Europe PMC	Janet University Network	193.62.193.83	Cambridge, UK
OSF Preprints	Google cloud	35.190.84.173	Missouri, US
Authorea Preprint Repository	Amazon	52.44.46.195 (no more used)	Virginia, US
	Cloudflare (updated)	2606:4700:4400::ac40:985d 104.18.35.163 172.64.152.93	Anycasted Anycasted Anycasted
Cambridge Open Engage	Cloudflare	104.17.50.120 104.17.51.120	Anycasted Anycasted
	(updated)	104.17.110.190 104.17.111.190	Anycasted Anycasted
Africarxiv	Google cloud	35.190.84.173	Anycasted
sci-hub.st	Ddos-guard	186.2.163.201	NL
sci-hub.se	Ddos-guard	186.2.163.219	NL
HAL	Renater	193.48.96.10	Hérault, FR
OAK	Blacknight Internet Solutions	46.22.140.125	IE

Table 7.2: Summary of open access directories

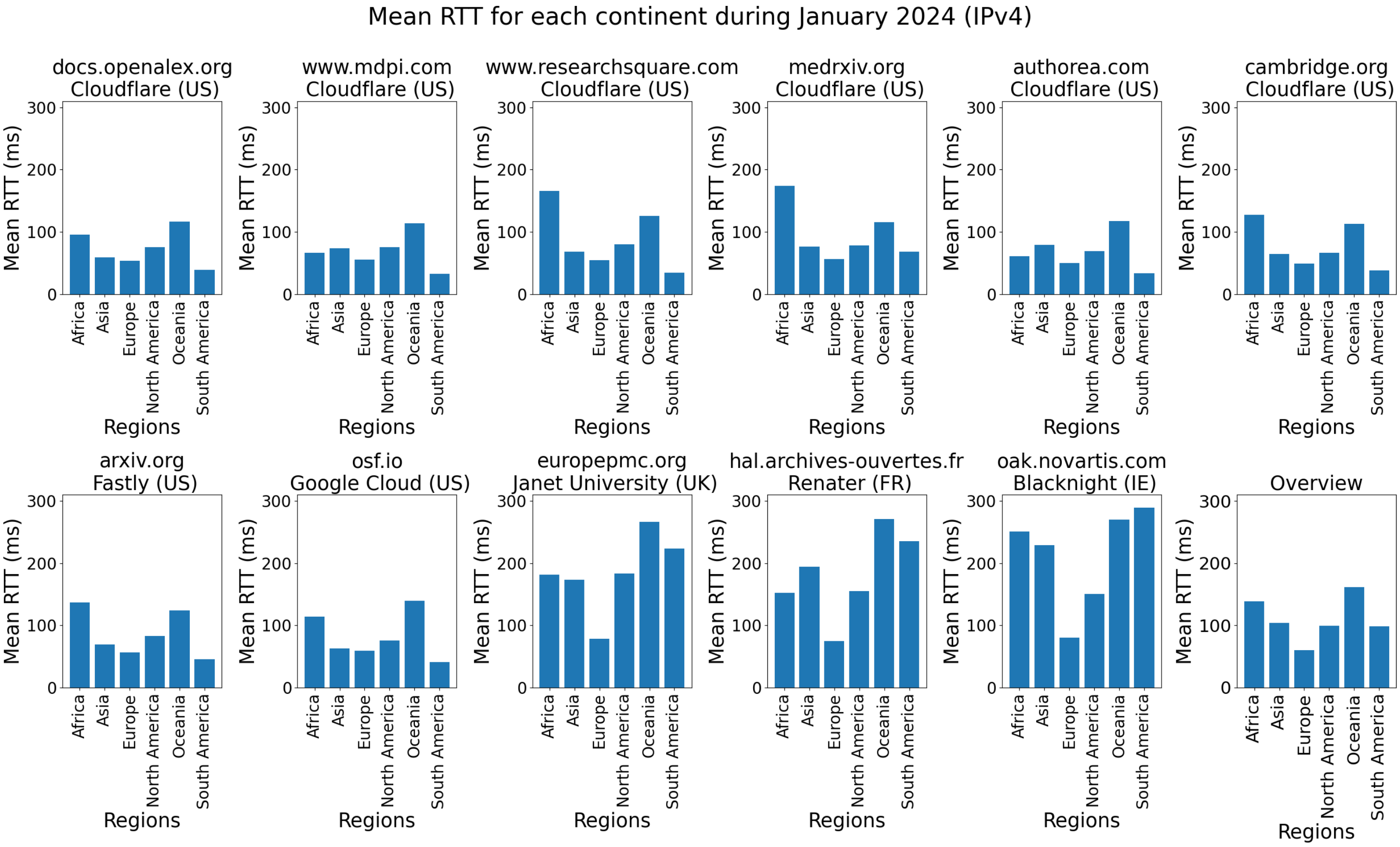

Figure 7.1: Mean RTT for each continent to all the repositories during January 2024 (IPv4).

Country	Fibre routes research (km)	Terrestrial fibres mapped (km)
South Africa	205,315	80,765
China	1,106,085	171,645
India	2,075,796	185,372
New Zealand	91,410	24,579
Argentina	86,816	79,150

Table 7.3: Summary of fibre routes research and terrestrial fibres mapped of the countries [55]

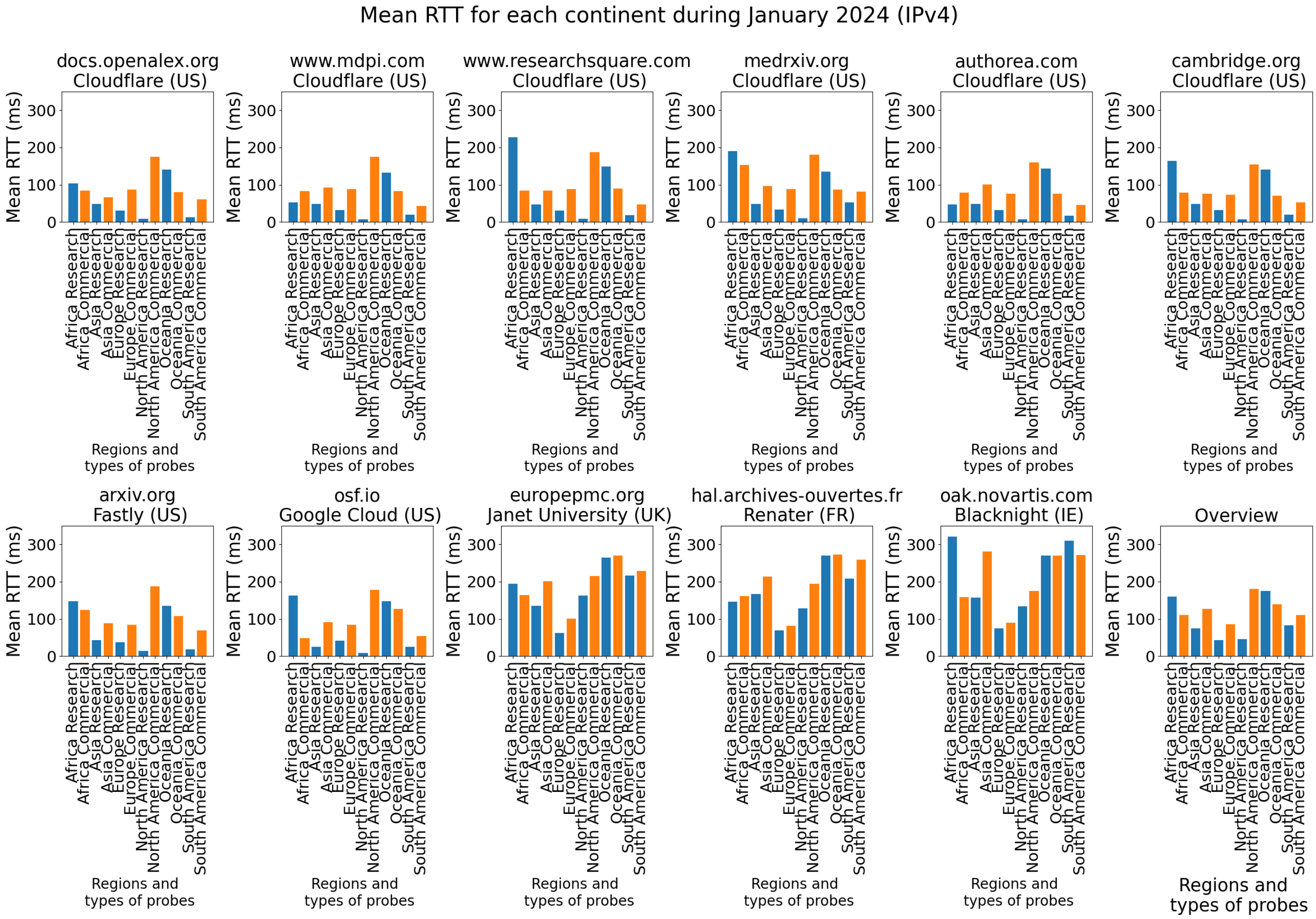

Figure 7.2: Comparison of mean RTT between commercial and research network for each continent to all repositories during January 2024 (IPv4)

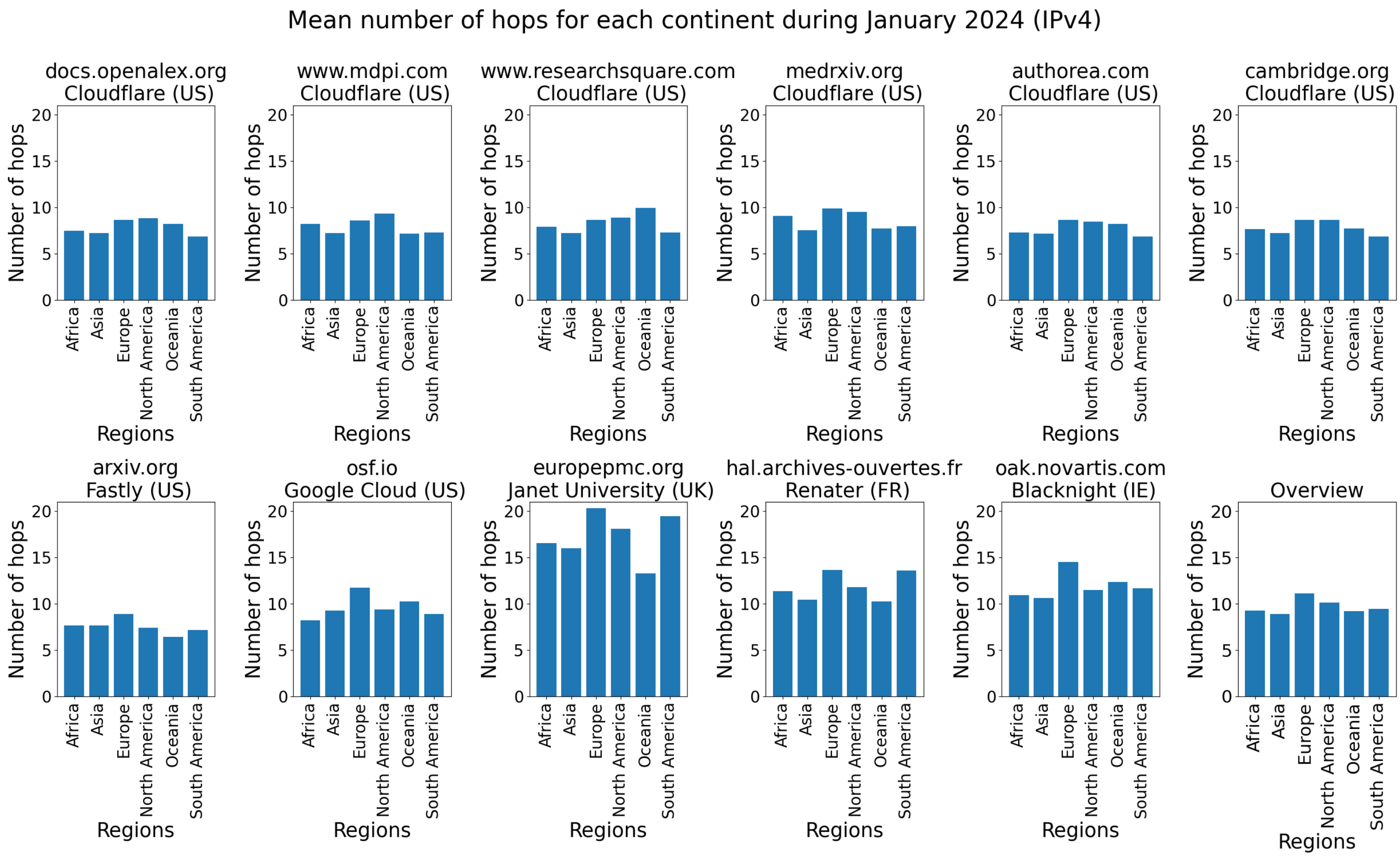

Figure 7.3: Mean number of hop (excluding measures that have not achieved their target) for each continent to all repositories during January 2024 (IPv4).

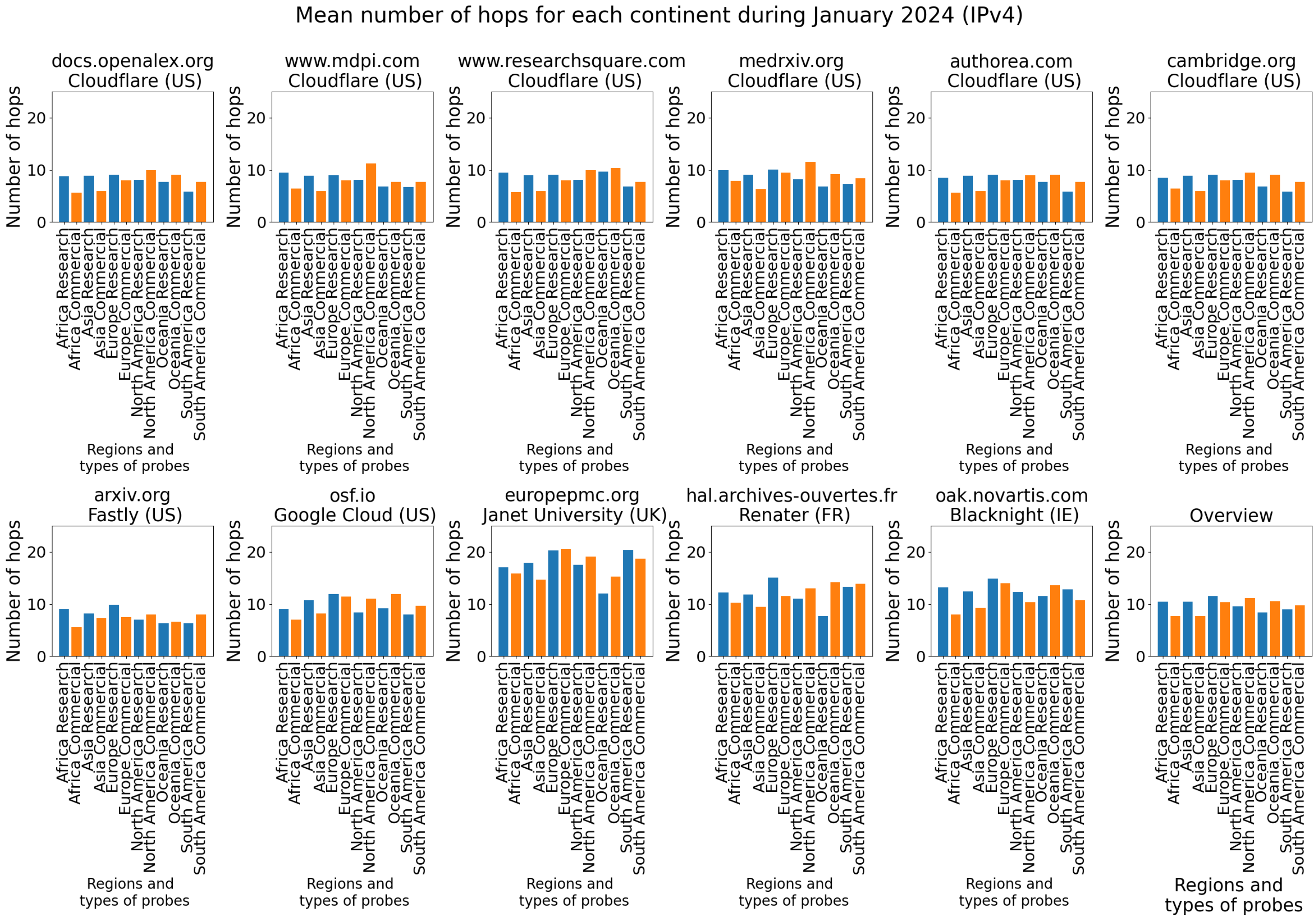

Figure 7.4: Comparison of mean number of hops between commercial and research network for each continent to all repositories during January 2024 (IPv4)

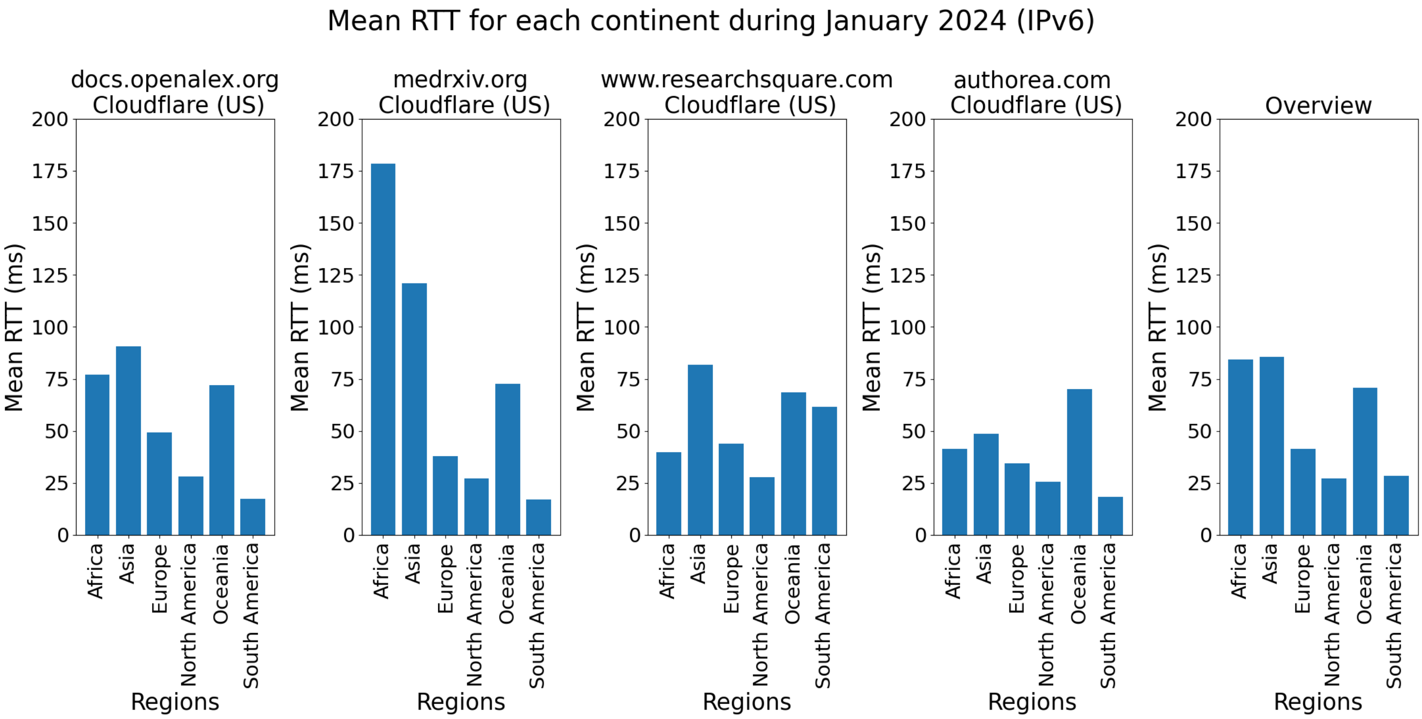

Figure 7.5: Mean RTT for each continent to all repositories supporting IPv6 during January 2024 (IPv6)

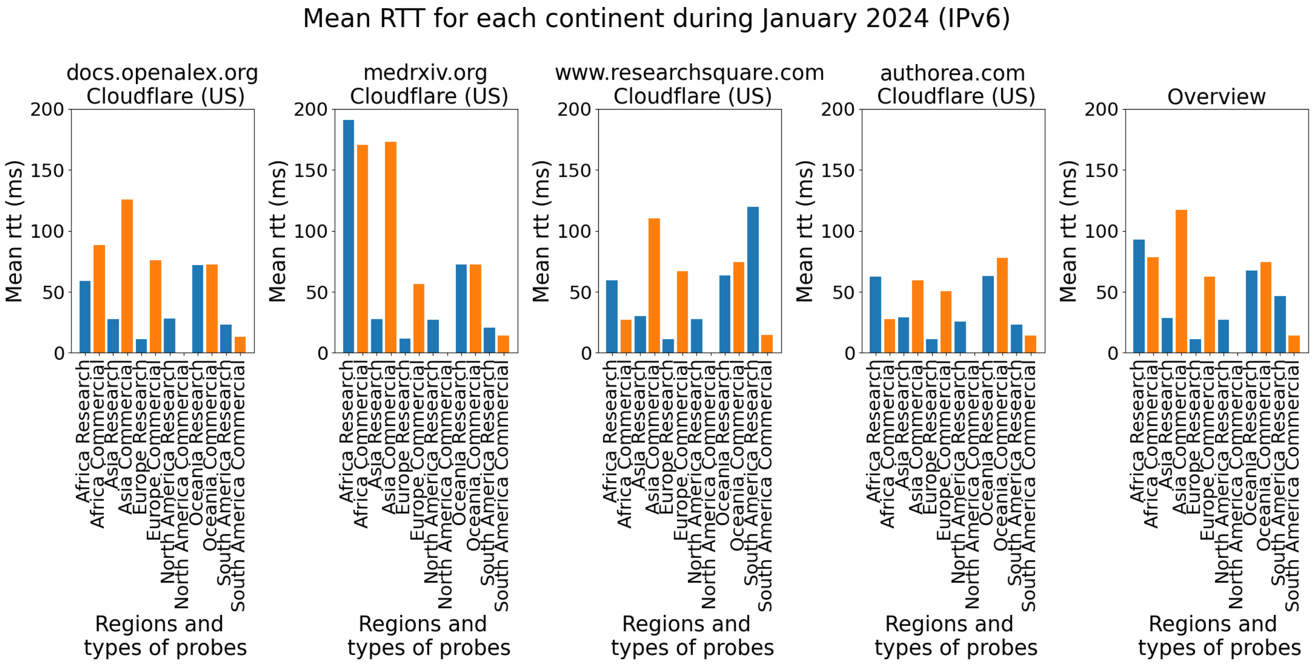

Figure 7.6: Comparison of mean RTT between commercial and research network for each continent to all directories supporting IPv6 during January 2024 (IPv6). There is no measurement from North America commercial probes.

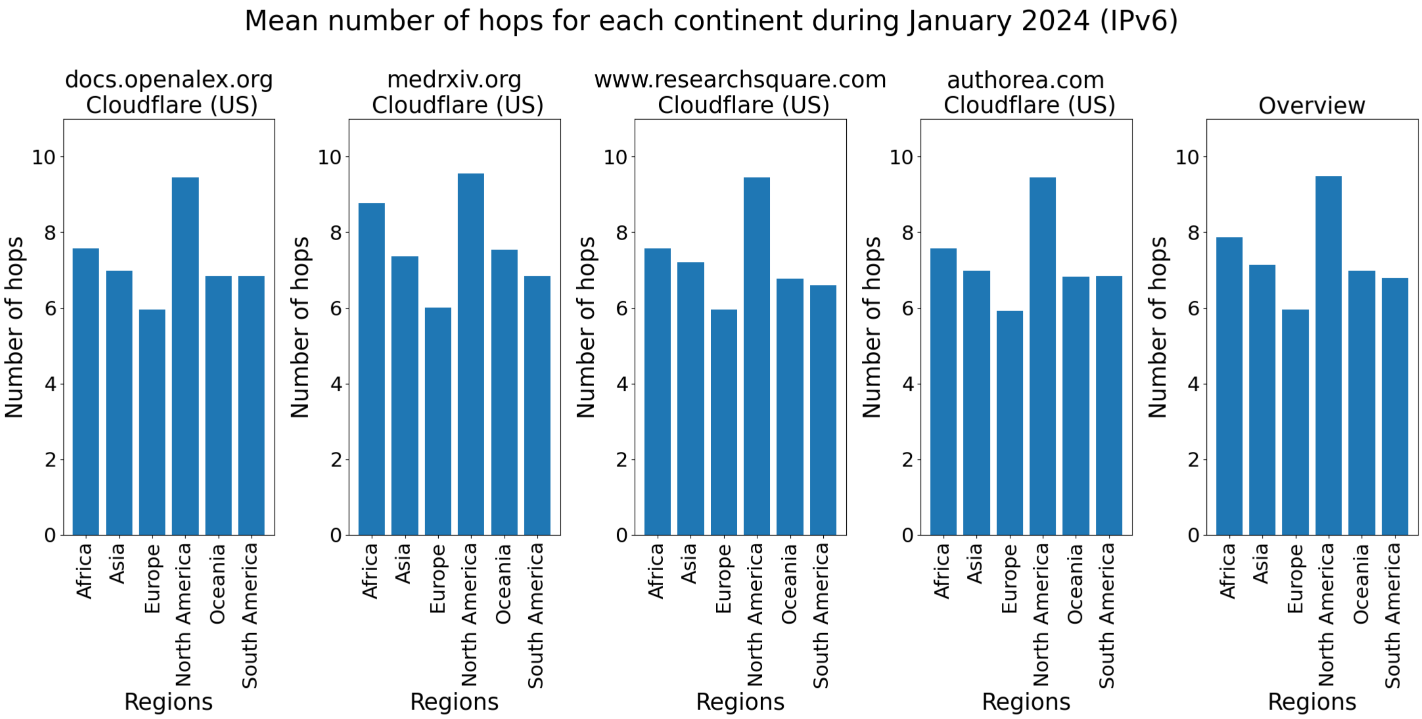

Figure 7.7: Mean number of hops for each continent to all repositories supporting IPv6 during January 2024 (IPv6).

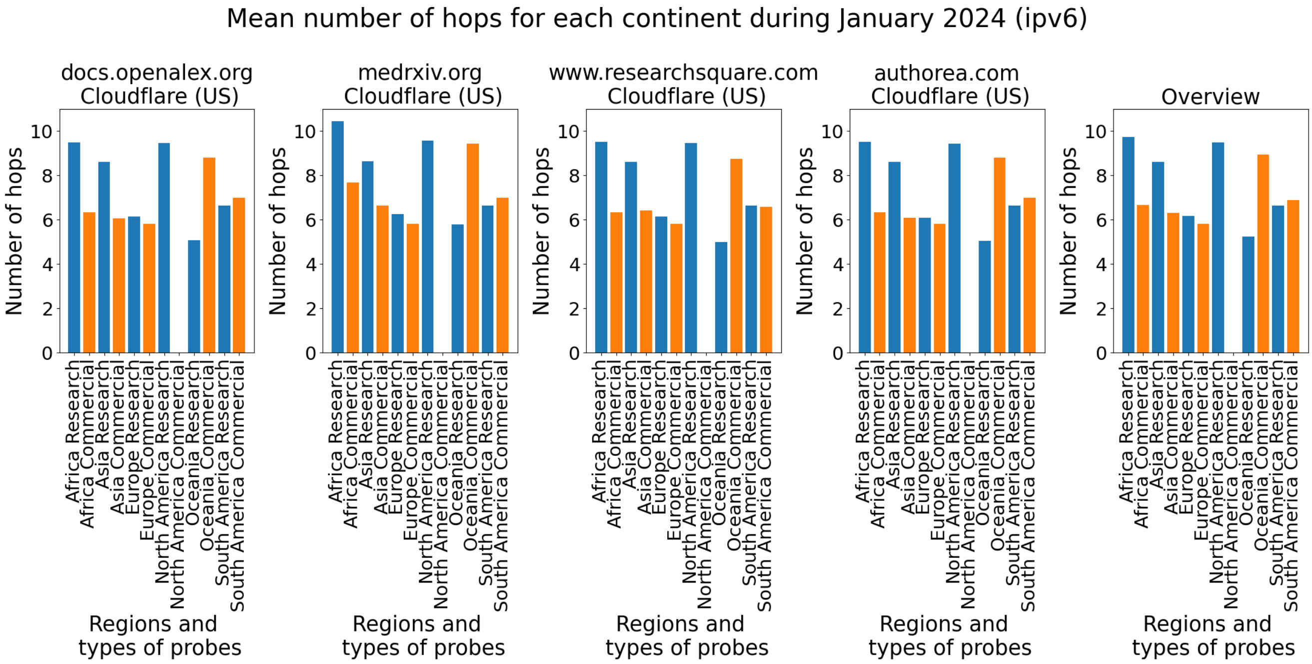

Figure 7.8: Comparison of mean number of hops between commercial and research network for each continent to all repositories supporting IPv6 during January 2024 (IPv6). There is no measurement from North America commercial probes.

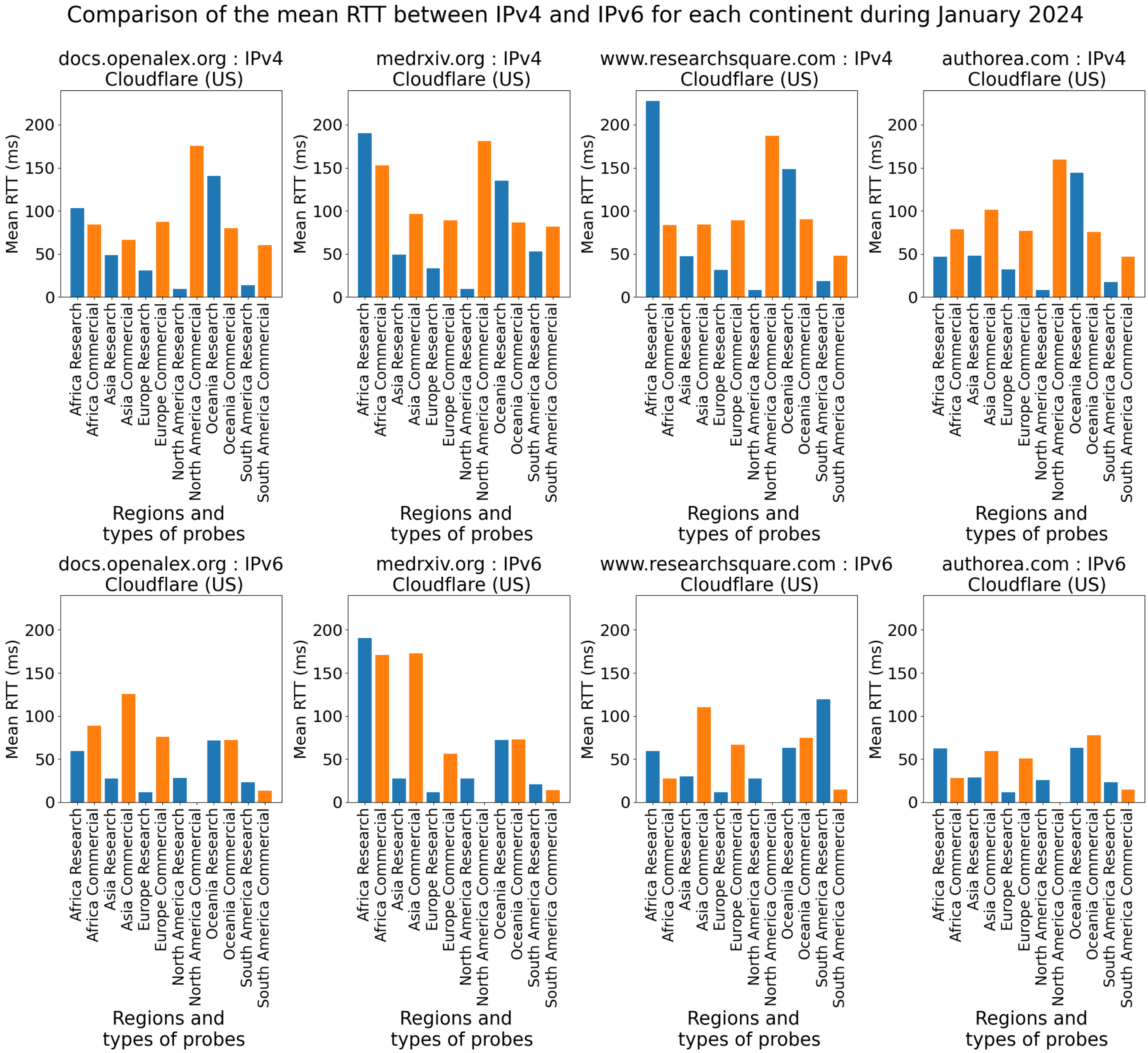

Figure 7.9: Comparison of mean RTT between IPv4 and IPv6 for each continent to the 4 repositories that support both IPv4 and IPv6 during January 2024. There are no measurements for North America Commercial in IPv6.

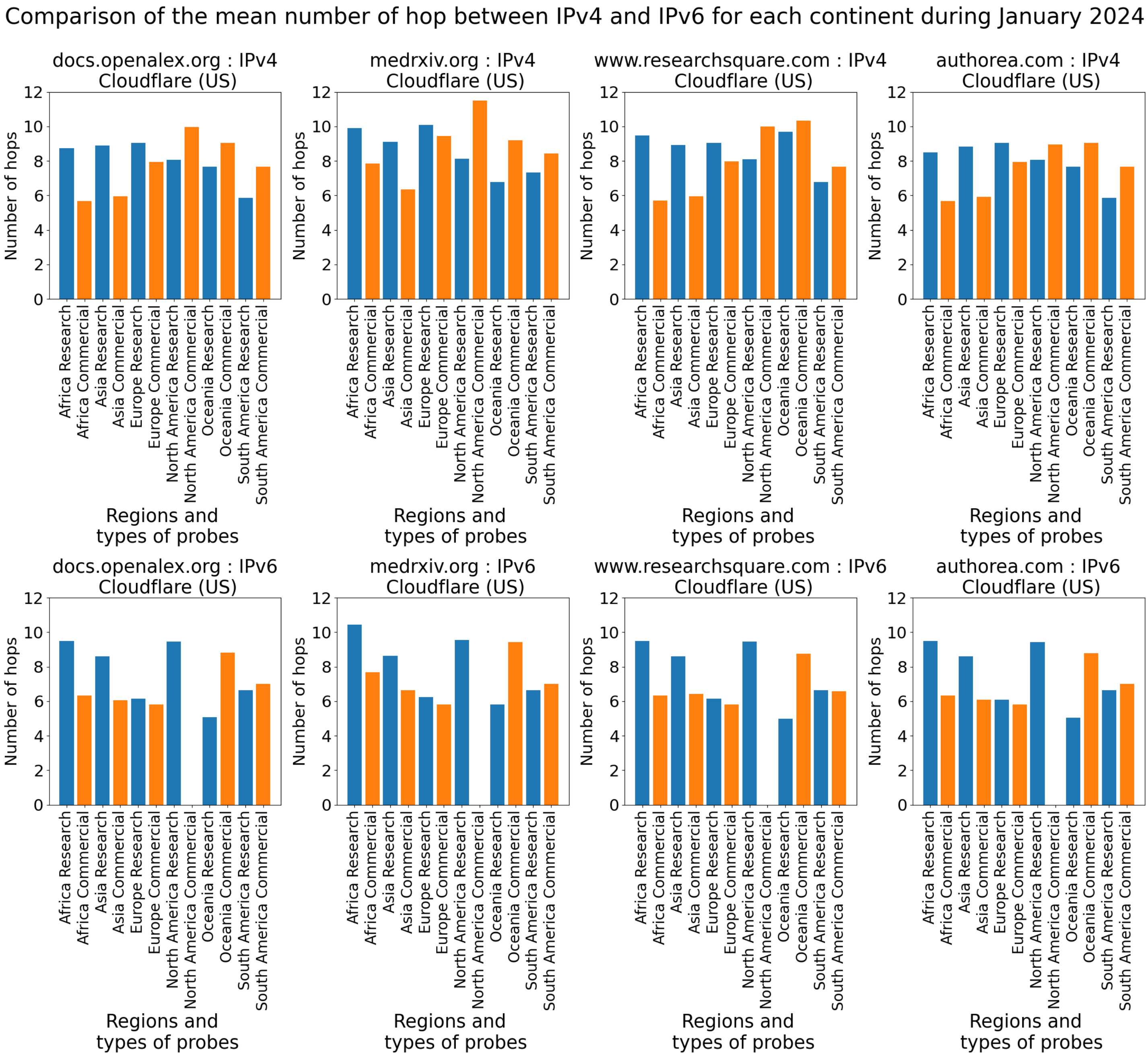

Figure 7.10: Comparison of the mean number of hops between IPv4 and IPv6 for each continent to the 4 repositories that support both IPv4 and IPv6 during January 2024. There are no measurements for North America Commercial in IPv6.

Bibliography

[1] Luci Abrahams et al. *Opening access to knowledge in Southern African universities*. SARUA, Pretoria, ZA, 2008.

[2] Berhan Oumer Adame. "The Ethiopian telecom industry: gaps and recommendations towards meaningful connectivity and a thriving digital ecosystem". In: *Heliyon* 7.10 (2021).

[3] Saqib Ali, R Les Cottrell, and Anjum Nveed. *Pinger malaysia-internet performance measuring project: A case study*. Tech. rep. SLAC National Accelerator Lab., Menlo Park, CA (United States), 2016.

[4] Saqib Ali et al. "Internet performance analysis of south asian countries using end-to-end internet performance measurements". In: *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*. IEEE. 2017, pp. 1319–1326.

[5] *anycasted.io*. last visited on 05-27-2024. URL: `https://anycasted.io/`.

[6] Kehinde Aruleba and Nobert Jere. "Exploring digital transforming challenges in rural areas of South Africa through a systematic review of empirical studies". In: *Scientific African* 16 (2022), e01190.

[7] Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwälder. "Lessons learned from using the ripe atlas platform for measurement research". In: *ACM SIGCOMM Computer Communication Review* 45.3 (2015), pp. 35–42.

[8] Vaibhav Bajpai et al. "Vantage point selection for IPv6 measurements: Benefits and limitations of RIPE Atlas tags". In: *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE. 2017, pp. 37–44.

[9] Tushar R Behera. *Internet Users in India: Statistics and Data (2024)*. last visited on 05-20-2024. 2024. URL: `https://www.grabon.in/indulge/tech/internet-users-statistics/`.

[10] Anthony Bergin and Sam Bashfield. *Digital age lies vulnerable to threats from underwater*. last visited on 05-25-2024. 2021. URL: `https://www.aspi.org.au/opinion/digital-age-lies-vulnerable-threats-underwater`.

[11] Theodore C Bergstrom et al. “Evaluating big deal journal bundles”. In: *Proceedings of the National Academy of Sciences* 111.26 (2014), pp. 9425–9430.

[12] Dominique Boullier. “Internet est maritime : les enjeux des câbles sous-marins”. In: *Revue internationale et stratégique 2014/3 (n° 95)*. IEEE. 2014, pp. 149–158.

[13] Anais Bozino. *Prix Liber : pour une science ouverte, transparente et gratuite*. last visited on 05-27-2024. 2020. URL: `https://www.static.inrae.fr/actualites/prix-liber-science-ouverte-transparente-gratuite`.

[14] Lane Burdette. “Leveraging Submarine Cables for Political Gain: US Responses to Chinese Strategy.” In: *Journal of Public & International Affairs* (2021).

[15] Matt Calder et al. “Analyzing the Performance of an Anycast CDN”. In: *Proceedings of the 2015 Internet Measurement Conference*. 2015, pp. 531–537.

[16] Esteban Carisimo et al. “A first look at the Latin American IXPs”. In: *ACM SIGCOMM Computer Communication Review* 50.1 (2020), pp. 18–24.

[17] Esteban Carisimo et al. “Studying the evolution of content providers in IPv4 and IPv6 internet cores”. In: *Computer Communications* 145 (2019), pp. 54–65.

[18] *cdnfinder*. 2024. URL: `https://www.cdnfinder.com/`.

[19] *CERNET*. last visited on 05-25-2024. URL: `https://www.cernet.edu.cn/`.

[20] Bert Chapman. “Undersea cables: The ultimate geopolitical chokepoint”. In: (2021).

[21] Josiah Chavula et al. “Quantifying the effects of circuitous routes on the latency of intra-Africa internet traffic: a study of research and education networks”. In: *e-Infrastructure and e-Services for Developing Countries: 6th International Conference, AFRICOMM 2014, Kampala, Uganda, November 24-25, 2014, Revised Selected Papers 6*. Springer. 2015, pp. 64–73.

[22] Marshini Chetty et al. “Measuring broadband performance in South Africa”. In: *Proceedings of the 4th Annual Symposium on Computing for Development*. 2013, pp. 1–10.

[23] China Internet Network Information Center (CNNIC). *The 52nd Statistical Report on China’s Internet Development*. Aug. 2023.

[24] Tobias LIEBETRAU Christian BUEGER and Jonas FRANKEN. *Security threats to undersea communications cables and infrastructure – consequences for the EU*. Tech. rep. European Parliament, 2016.

[25] Cisco. *Cisco Annual Internet Report (2018–2023) White Paper.* last visited on 05-25-2024. 2020. URL: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[26] Cloudflare. *What is a data center?* last visited on 05-27-2024. URL: https://www.cloudflare.com/learning/cdn/glossary/data-center/.

[27] Lorenzo Corneo et al. "Surrounded by the clouds: A comprehensive cloud reachability study". In: *Proceedings of the Web Conference 2021.* 2021, pp. 295–304.

[28] Jakub Czyz et al. "Measuring ipv6 adoption". In: *Proceedings of the 2014 ACM Conference on SIGCOMM.* 2014, pp. 87–98.

[29] Rami Al-Dalky, Michael Rabinovich, and Kyle Schomp. "A Look at the ECS Behavior of DNS Resolvers". In: *Proceedings of the Internet Measurement Conference.* 2019, pp. 116–129.

[30] The Khang Dang et al. "Cloudy with a chance of short rtts: Analyzing cloud connectivity in the internet". In: *Proceedings of the 21st ACM Internet Measurement Conference.* 2021, pp. 62–79.

[31] Amogh Dhamdhere et al. "Measuring the deployment of IPv6: topology, routing and performance". In: *Proceedings of the 2012 Internet Measurement Conference.* 2012, pp. 537–550.

[32] Strickland Eliza. *Why the Japan Earthquake Didn't Take Down the Country's Internet.* last visited on 05-25-2024. 2011. URL: https://spectrum.ieee.org/why-the-japan-earthquake-didnt-cripple-the-countrys-internet.

[33] EUA. *The new university Open Access checklist, A practical guide on implementation.* 2021.

[34] European Union. *Horizon 2020.* last visited on 05-27-2024. URL: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en.

[35] Rodérick Fanou, Pierre Francois, and Emile Aben. "On the diversity of interdomain routing in africa". In: *Passive and Active Measurement: 16th International Conference, PAM 2015, New York, NY, USA, March 19-20, 2015, Proceedings 16.* Springer. 2015, pp. 41–54.

[36] Rodérick Fanou, Francisco Valera, and Amogh Dhamdhere. "Investigating the Causes of Congestion on the African IXP substrate". In: *Proceedings of the 2017 Internet Measurement Conference.* 2017, pp. 57–63.

[37] Rodérick Fanou et al. "Pushing the frontier: Exploring the african web ecosystem". In: *Proceedings of the 25th International Conference on World Wide Web*. 2016, pp. 435–445.

[38] Rodérick Fanou et al. "Unintended consequences: Effects of submarine cable deployment on Internet routing". In: *Passive and Active Measurement: 21st International Conference, PAM 2020, Eugene, Oregon, USA, March 30–31, 2020, Proceedings 21*. Springer. 2020, pp. 211–227.

[39] Agustin Formoso et al. "Deep diving into africa's inter-country latencies". In: *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE. 2018, pp. 2231–2239.

[40] Hernán Galperín. "Localizing internet infrastructure: Cooperative peering in latin america". In: *Telematics and Informatics* 33.2 (2016), pp. 631–640.

[41] *GEANT*. last visited on 05-25-2024. URL: `https://geant.org/`.

[42] SB Ghosh and Anup Kumar Das. "Open access and institutional repositories—a developing country perspective: A case study of India". In: *IFLA journal* 33.3 (2007), pp. 229–250.

[43] Forough Golkar, Thomas Dreibholz, and Amund Kvalbein. "Measuring and comparing internet path stability in ipv4 and ipv6". In: *2014 International Conference and Workshop on the Network of the Future (NOF)*. IEEE. 2014, pp. 1–5.

[44] UC Berkeley Gretchen Kell. *Why UC split with publishing giant Elsevier*. last visited on 05-27-2024. 2019. URL: `https://www.universityofcalifornia.edu/news/why-uc-split-publishing-giant-elsevier`.

[45] *HAL*. last visited on 04-04-2024. 2024. URL: `https://doc.hal.science/`.

[46] Thomas Holterbach et al. "Quantifying interference between measurements on the RIPE Atlas platform". In: *Proceedings of the 2015 Internet Measurement Conference*. 2015, pp. 437–443.

[47] Mu-Hsuan Huang and Mei-Jhen Huang. "An analysis of global research funding from subject field and funding agencies perspectives in the G9 countries". In: *Scientometrics* 115.2 (2018), pp. 833–847.

[48] Qi Huang et al. "An analysis of Facebook photo caching". In: *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles*. 2013, pp. 167–181.

[49] indec. *Tecnología*. last visited on 05-20-2024. 2023. URL: `https://www.indec.gob.ar/indec/web/Nivel3-Tema-4-26#:~:text=En%20el%20cuarto%20trimestre%20de,sea%20red%20fija%20o%20m%C3%B3vil)`..

[50] Telecom Regulatory Authority of India. *Annual report 2022-2023*. last visited on 05-20-2024. 2023. URL: https://trai.gov.in/sites/default/files/Annual_Report_20022024_0.pdf.

[51] Classe Internationale. *Les câbles sous-marins au centre des enjeux de puissance mondiale : EllaLink, une alternative de gouvernance de l'Internet*. last visited on 05-25-2024. 2021. URL: https://classe-internationale.com/2021/06/15/les-cables-sous-marins-au-centre-des-enjeux-de-puissance-mondiale-ellalink-une-alternative-de-gouvernance-de-linternet/.

[52] *INTERNET2*. last visited on 05-25-2024. URL: https://internet2.edu/.

[53] *internet2 : network*. last visited on 05-20-2024. 2024. URL: https://internet2.edu/network/.

[54] Internetnz. *New Zealand's Internet insights 2020*. last visited on 05-20-2024. 2020. URL: https://internetnz.nz/new-zealands-internet-insights/new-zealands-internet-insights-2020/.

[55] ITU. *Infrastructure Connectivity Map*. last visited on 05-20-2024. 2024. URL: https://bbmaps.itu.int/bbmaps/.

[56] ITU. *Measuring digital development Facts and Figures 2021*. last visited on 05-25-2024. 2021. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf.

[57] ITU. *Measuring digital development Facts and Figures 2023*. last visited on 05-25-2024. 2023. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2023.pdf.

[58] Kirk L. Johnson et al. "The measured performance of content distribution networks". In: *Computer Communications* 24.2 (2001), pp. 202–206.

[59] Jean-Baptiste Jouffray et al. "The blue acceleration: the trajectory of human expansion into the ocean". In: *One Earth* 2.1 (2020), pp. 43–54.

[60] Lisa Lasselin. *Publications scientifiques : à Lille comme ailleurs on paie le prix fort*. last visited on 05-27-2024. 2021. URL: https://www.lavoixdunord.fr/1120519/article/2021-12-29/publications-scientifiques-lille-comme-ailleurs-paie-le-prix-fort.

[61] Fuliang Li et al. "A case study of ipv6 network performance: Packet delay, loss, and reordering". In: *Mathematical Problems in Engineering* 2017 (2017).

[62] Ying Liu. "Current situation and prospect of CERNET". In: *China's e-Science Blue Book 2020* (2021), pp. 327–334.

[63] Seyed Ali Mirheidari et al. "Web cache deception escalates!" In: *31st USENIX Security Symposium (USENIX Security 22)*. 2022, pp. 179–196.

[64] *Netstage*. last visited on 05-25-2024. URL: https://www.netsage.global/.

[65] Charles R Newton. "Research and open access from low-and middle-income countries". In: *Developmental medicine and child neurology* 62.5 (2020), p. 537.

[66] Rosanna Norman. *BELLA: EllaLink cable gets go-ahead*. last visited on 05-25-2024. 2021. URL: https://connect.geant.org/2019/01/08/bella-ellalink-cable-gets-go-ahead.

[67] Alan C O'Connor et al. "Economic Impacts of Submarine Fiber Optic Cables and Broadband Connectivity in Indonesia". In: *Research Triangle Park, NC, USA: RTI International* (2020).

[68] *Orient Plus*. last visited on 05-25-2024. URL: https://orientplus.eu/.

[69] PCH. *Internet Exchange Directory*. last visited on 05-25-2024. 2024. URL: https://www.pch.net/ixp/dir#!mt-sort=reg%2Casc!mt-zoom=%5B1.6084981600629784%2C-0.1565899944628556%2C-0.3319838231541427%5D.

[70] PCH. *Internet Exchange Point Growth by Region*. last visited on 05-25-2024. 2024. URL: https://www.pch.net/ixp/summary_growth_by_region.

[71] Matthew Prince. *A Brief Primer on Anycast*. last visited on 05-27-2024. 2011. URL: https://blog.cloudflare.com/a-brief-anycast-primer/.

[72] *Reannz live weathermap*. last visited on 06-21-2024. URL: https://weathermap.reannz.co.nz/index.php?src=and31.

[73] *RIPE Atlas population coverage (IPv4)*. last visited on 05-27-2024. URL: https://sg-pub.ripe.net/petros/population_coverage/map.html?version=4&public_only=false.

[74] *RIPE Atlas population coverage (IPv6)*. last visited on 05-27-2024. URL: https://sg-pub.ripe.net/petros/population_coverage/map.html?version=6&public_only=false.

[75] Carlos Rodrigues. *cdnfinder*. last visited on 05-20-2024. 2024. URL: https://blog.cloudflare.com/fr-fr/ipv6-from-dns-pov-fr-fr/.

[76] Romeo Agan Salac and Yun Seon Kim. "A study on the internet connectivity in the Philippines". In: *Asia Pacific Journal of Business Review* 1.1 (2016), pp. 67–88.

[77] *SANReN*. last visited on 05-25-2024. URL: https://www.sanren.ac.za/south-african-nren/.

[78] Antoine Saverimoutou, Bertrand Mathieu, and Sandrine Vaton. “Influence of internet protocols and CDN on web browsing”. In: *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE. 2019, pp. 1–5.

[79] Pavlos Sermpezis et al. “Bias in Internet Measurement Platforms”. In: *2023 7th Network Traffic Measurement and Analysis Conference (TMA)*. IEEE. 2023, pp. 1–10.

[80] KEMP Simon. *DIGITAL 2023: NEW ZEALAND*. last visited on 05-20-2024. 2023. URL: https://datareportal.com/reports/digital-2023-new-zealand.

[81] Stephen D Strowes. “Diurnal and weekly cycles in IPv6 Traffic”. In: *Proceedings of the 2016 Applied Networking Research Workshop*. 2016, pp. 65–67.

[82] Alma Swan et al. *Policy guidelines for the development and promotion of open access*. UNESCO Publishing, 2012.

[83] TeleGeography. *Cloud Infrastructure Map*. last visited on 05-20-2024. 2024. URL: https://www.cloudinfrastructuremap.com/#/service/cloud-regions.

[84] *TENET*. last visited on 05-25-2024. URL: https://www.tenet.ac.za/.

[85] *Twelve99 Looking Glass*. last visited on 05-27-2024. URL: https://lg.twelve99.net/.

[86] *UBUNTUNET ALLIANCE*. last visited on 05-25-2024. URL: https://ubuntunet.net/.

[87] Université de Liège. *Chiffres-clés de l’Université de Liège*. last visited on 05-27-2024. URL: https://www.uliege.be/cms/c_9038278/fr/presentation-longue-chiffres-cles.

[88] Université de Lorraine. *Un point sur l’accès aux publications scientifiques à l’Université de Lorraine*. last visited on 05-27-2024. 2019. URL: https://factuel.univ-lorraine.fr/node/11065.

[89] Université Libre de Bruxelles. *Chiffres-clés de l’Université Libre de Bruxelles*. last visited on 05-27-2024. URL: https://www.ulb.be/medias/fichier/ppt-institutionnel-ulb-dec-2022-fr_1671546594041-pdf.

[90] Xianwen Wang et al. “Science funding and research output: a study on 10 countries”. In: *Scientometrics* 91.2 (2012), pp. 591–599.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**