Multi-Layered Telemetry Assessing Global Performance of LEO Internet Providers

Towards a Global Telemetry System for Evaluating LEO ISP Performance

Vlad-Ștefan Graure[1]

Supervisor(s): Nitinder Mohan[1], Tanya Shreedhar[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Vlad-Ștefan Graure
Final project course: CSE3000 Research Project
Thesis committee: Nitinder Mohan, Tanya Shreedhar, Qing Wang

Abstract

The rise of Low-Earth-Orbit (LEO) satellite networks, such as Starlink, has transformed global connectivity, enabling high-speed internet access in previously underserved regions. However, existing research lacks a unified framework to evaluate and compare the performance of LEO ISPs against terrestrial alternatives using heterogeneous measurement datasets. In this work, we present a methodology for harmonizing and standardizing passive internet measurements from M-Lab's NDT7 and Cloudflare's AIM datasets, implemented in the form of the **Global Telemetry System**. These sources are integrated through schema unification, filtering, and normalization to produce a reproducible and geographically comprehensive telemetry dataset. We introduce a server-based filtering approach to mitigate geographic and routing biases, and we evaluate aggregation methods to align measurement distributions across datasets. Our results demonstrate the dataset integration methodology preserves key distributional properties, enabling fair and statistically consistent merging of measurements from the two sources. This work represents a first step toward a scalable and extensible telemetry infrastructure for assessing next-generation global internet services.

1 Introduction

Over the past few years, the Internet has undergone significant evolution, with widespread expansion into remote regions made possible through the deployment of new fiber-optic cables and advancements in telecommunication technologies [11, 17]. The emergence of Low-Earth-Orbit (LEO) Internet Service Providers—such as Starlink [25], Amazon Kuiper [1], and OneWeb [19] has further revolutionized connectivity, enabling Internet access even in some of the most isolated areas on Earth, including remote islands, mountainous regions, and vast deserts. LEO networks are composed of large constellations of satellites orbiting between 160 and 2,000 km above Earth's surface, promising to offer not only global coverage but also significantly reduced latency, thereby making reliable, high-speed internet accessible in even the most challenging environments.

The Starlink service developed by SpaceX stands out among Low-Earth-Orbit (LEO) Internet Service Providers (ISPs) due to its deployment of over 63,000 satellites [21], which serve more than 5 million users across 125 countries [23]. Despite the immense potential that Starlink could unlock in the global connectivity landscape, only a limited number of studies have explored this area in depth. Most studies on the Starlink network primarily focused on analyzing its network topology and measuring its performance [15, 20, 26, 30]. These studies rely on passive telemetry data, obtained from sources like Cloudflare AIM [4] and M-Lab [12], or on active measurements using RIPE Atlas probes [16] or collected directly by the research teams. However, a key limitation of these studies is the lack of a global telemetry system, which restricts the ability to obtain a comprehensive view of the Starlink network's performance on a global scale and requires the researchers to combine the different data sources themselves. A global telemetry system would not only allow for real-time monitoring across regions but also enable more accurate comparisons between Starlink and terrestrial ISPs. Such

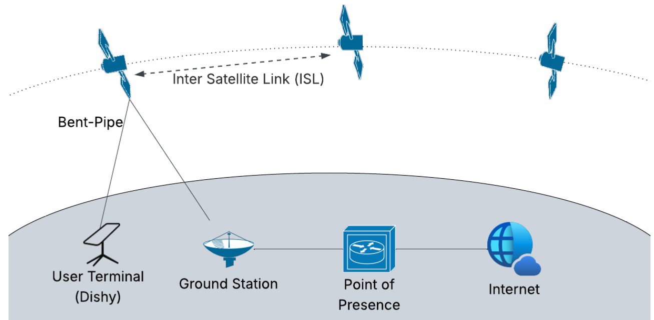


Figure 1: Bent-pipe architecture of the Starlink system. One or more satellites, connected via inter-satellite links (ISLs), establish the connection between the user terminal and the ground station.

comparisons are essential for understanding how LEO satellite networks perform relative to traditional terrestrial infrastructure in terms of latency, coverage, and overall service quality.

This paper aims to develop a global internet telemetry system by integrating passive measurements from heterogeneous sources, enabling fair and transparent comparisons between LEO satellite networks and terrestrial ISPs. Specifically, we evaluate the methodological compatibility of M-Lab's NDT7 dataset with Cloudflare's AIM data and propose techniques for normalizing, integrating, and debiasing the combined datasets. We design and implement a system that continuously harmonizes these measurements into a unified telemetry dataset. The resulting system provides passive, scalable, and reproducible telemetry that supports researchers and policymakers in accurately assessing and comparing internet performance across diverse access technologies. To ensure its longevity and continued impact, the core system is released as an open-source CLI tool [28] and has been integrated into the LEO Viewer project [9] for continuous deployment and data collection, enabling sustained use and further development by the community.

The remainder of this paper is organized as follows. Section 2 provides an overview of Starlink's architecture and the telemetry data sources used in our study. Section 3 outlines the methodology for selecting relevant ISPs, designing a unified schema, preprocessing measurements, and mitigating server-induced bias. Section 4 details the implementation of the integration pipeline that automates the harmonization of NDT7 and AIM datasets. Section 5 presents the evaluation results, including distribution comparisons and the impact of server-based filtering. Finally, Section 6 concludes the paper with reflections on limitations and outlines directions for future work.

2 Background

Starlink is a LEO ISP operated by SpaceX, having approximately 6,750 operational satellites [24]. The satellites operate on three different orbits: 53°, 70°, and 97.6°, with the majority operating in the 53° [15]. A "bent-pipe" architecture is employed by the ISP, as illustrated in Figure 1. The satellite the user terminal connects to might relay the data through other satellites, forming an "extended bent-pipe" before reaching the Ground Station (GS). Evidence was found that a global scheduler allocates the user-satellite-GS connection every 15 seconds, selecting the satellites based on several factors, such as inclination or whether they are being sunlit [15, 26].

NDT7 (Network Diagnostic Tool 7) [12] is a protocol employed by M-Lab in their Internet speed test to assess application-level upload and download performance using WebSockets over TLS. The test establishes a secure WebSocket connection over a single TCP connection configured with the BBR congestion control algorithm. During a fixed 10-second interval, the client uploads or downloads as much data as possible, thereby measuring the achievable goodput under current network conditions. The test is user-initiated and typically connects to the geographically closest available server, which is either operated by M-Lab or hosted on Google Cloud infrastructure. M-Lab servers are usually located in large data centers [13]. The resulting measurements, reported by both the client and the server, are made publicly available via M-Lab's BigQuery dataset. Each test includes multiple snapshots, capturing performance metrics at different moments throughout the session.

Cloudflare AIM (Aggregated Internet Measurements) is an initiative designed to provide end-users with a comprehensive understanding of their internet performance. This is achieved through the provision of AIM scores, which evaluate how good the user's connection supports various tasks such as gaming, streaming, and video conferencing. Each category receives a label based on the results of Cloudflare Radar's internet speed test: "Great", "Good", "Average", "Poor", "Bad", or "Unknown". The speed test evaluates several key metrics, including download and upload throughput, loaded and unloaded latency and jitter, and packet loss. Throughput tests are conducted over HTTPS using HTTP/1.1 over a single TCP BBR connection. The tests involve progressively larger files being downloaded or uploaded to measure the achievable throughput. Latency and jitter are assessed using HTTP-based methods, while packet loss is measured via UDP using WebRTC, which is used to simulate real-time communication such as video calling. A detailed method for aggregating these measurements into the AIM scores is available on Cloudflare Radar's website [27]. The test is automatically initiated when the user visits the test's webpage. Typically, the server chosen is geographically closest to the user and part of Cloudflare's CDN network. The resulting data is publicly accessible via BigQuery thanks to the collaboration between M-Lab and Cloudflare Radar.

We select NDT7 and Cloudflare AIM as our primary measurement sources due to their public accessibility and global coverage. NDT7, maintained by M-Lab, provides high-frequency, application-level throughput data using a standardized test protocol and is widely used in academic studies on internet performance. Its data is open, well-documented, and continuously updated via Google BigQuery. Cloudflare AIM complements NDT7 by offering a broader suite of real-world performance metrics, directly relevant to user experience in common applications such as gaming and video conferencing. Its focus on aggregated user experience metrics provides a valuable contrast to NDT7's raw measurements.

3 Methodology

3.1 Identifying Top ISPs per Country

To assess the performance of terrestrial versus LEO (Low Earth Orbit) ISPs, we first identify the top terrestrial ISPs for each country. We focus on the top five ISPs per country to ensure adequate representation, especially in larger countries such as the United States or Brazil. Smaller countries, such as island nations, may not have five distinct ISPs, but the limit ensures consistency and scalability across the dataset.

Each Internet Service Provider is associated with an Autonomous System Number (ASN). For instance, Starlink operates under *AS14593*, while OneWeb uses *AS800*. To determine the top ISPs per country, we leverage data from the CAIDA AS Rank project, developed by the Center for Applied Internet Data Analysis (CAIDA) at the University of California, San Diego. This project ranks Autonomous Systems (ASes) globally based on their customer cone size, which represents the total number of ASes that are routed through a particular AS, including the AS itself, its direct customers, and all downstream customer relationships. These relationships between different ISPs are inferred from BGP path data.

The AS Rank data is accessible via CAIDA's GraphQL API [3], and the schema is documented in detail on their official website [2]. For our analysis, we download data about all ISPs and import it into a local PostgreSQL database, enabling efficient execution of complex queries. We develop a query that ranks ASes within each country by their global rank and selects the top five ASes where available. The resulting dataset, *Top Five ASNs per Country*, is publicly available in our GitHub repository [8].

In parallel, we create a dataset of countries where Starlink measurements are available. We identify these countries by performing a union of countries from Cloudflare's `speedtest_speed1` and M-Lab's `NDT7` datasets, where Starlink-related measurements (identified via ASN *AS14593*) have been recorded since January 1, 2023. The queries are present in Appendix A: Listing 1 is used to identify the countries with Starlink measurements, while Listing 2 identifies the top five ISPs from countries with Starlink measurements. This filtered subset of countries is then used in conjunction with the CAIDA-based ranking to extract the top five terrestrial ISPs in regions where Starlink is active.

3.2 Designing the Schema

We begin by surveying prior research efforts focused on measuring Starlink's network performance, with the goal of identifying the key metrics a telemetry system should capture. Pan et al. [20] leveraged latency and ping measurements to evaluate Starlink's performance and to explore the structure and operational characteristics of the network. Tanveer et al. [26] employed latency and packet loss rates to investigate Starlink's internal scheduling and behavior. Mohan et al. [15] utilized upload and download throughput, latency, jitter, and packet loss rate to assess Starlink's performance across common real-world applications such as online gaming, video streaming, and conferencing.

To construct the common schema for the global telemetry system, we analyze the structure of the Cloudflare AIM and NDT7 datasets, examine how fields can be mapped between them, and identify the normalization and transformation techniques required to ensure data consistency.

3.2.1 Cloudflare AIM. The Cloudflare AIM dataset provides both aggregated AIM scores and the raw measurement data from which these scores are derived. For our analysis, we focus exclusively on the raw measurements, which offer a more granular and interpretable view of network performance. As mentioned in Section 2, Cloudflare Radar's Internet speed test performs multiple independent measurements per session. Consequently, the dataset schema is not entirely flat—some metrics, such as throughput and latency, are stored as arrays with one value per test, while others are reported as single scalar values. This is either because only a single test was performed for that metric (e.g., packet loss), or because the metric is computed across all available tests (e.g., jitter is derived from latency measurements).

In line with previous network measurement studies, we focus on a subset of metrics that are most commonly used to evaluate key aspects of Internet performance. Our primary interest lies in the *upload* and *download* records, which provide arrays of transferred file sizes (in bytes) and their corresponding throughput values (in bits per second). These metrics are essential for assessing bandwidth capacity and performance consistency. To capture network responsiveness under load, we utilize the *loadedLatencyMs* field, which contains two arrays—one for upload and one for download—representing the round-trip latencies (in milliseconds) during active data transfer. These values reflect delay experienced during realistic usage conditions and are critical for latency-sensitive applications. We also incorporate the *loadedJitterMs* field, which reflects the variability in loaded latency. Additionally, we include the *packetLoss.lossRatio* value, a scalar indicating the proportion of packets lost during transmission.

Finally, to contextualize the performance data, we leverage client-side metadata including the autonomous system number (*clientASN*), geographic location (*clientCity*, *clientRegion*, *clientCountry*), and the time of the measurement (*measurementTime*, in UTC). For the server-side context in the Cloudflare AIM dataset, we use the *serverPoP* field, which is presented as an IATA airport code, indicating an approximate location of the server.

3.2.2 NDT7. In the NDT7 dataset, upload and download speed tests are conducted independently, meaning that only one of the *Upload* and *Download* records is populated per test. Both records share an identical schema, with the only difference being the direction of the test. During each M-Lab speed test, the client and server periodically exchange status updates, which are captured in the final dataset under the *raw* record. This record contains both *ServerMeasurements* and *ClientMeasurements*. To maintain alignment with Cloudflare's measurement approach, our analysis focuses on server measurements, which includes four key sub-records: *AppInfo*, *BBRInfo*, *TCPInfo*, and *ConnectionInfo*.

Our primary focus is on the *TCPInfo* record, which contains low-level TCP metrics collected by the server's kernel. Several of these fields correspond directly to those used in the Cloudflare AIM dataset, enabling cross-comparison and integration. Among the most important are the round-trip time (*RTT*) and its variance (*RTTVar*), both measured in microseconds. We interpret *RTT* as the network loaded latency and *RTTVar* as the jitter—i.e., the variability in round-trip delay—consistent with common usage in performance analysis.

The throughput and packet loss rate measured during the test are stored in the record *a* under the attributes *MeanThroughputMbps* and *LossRate*, respectively. Based on the type of test performed, which is determined by the presence of a non-null value in the corresponding record (i.e., upload or download), we populate the Global Telemetry System's upload or download throughput accordingly.

Beyond performance metrics, we also extract contextual metadata about both the client's and the server's network and location. Specifically, we use the UTC timestamp from the `a` record, the `Client.Geo` field to determine the client's city, region, and ISO 3166-1 alpha-2 country code, and the `Client.Network` field to identify the client's autonomous system number (ASN). Similarly, we leverage the `Server.Geo` field to obtain the server's city and country.

3.2.3 Schema Design. The final schema design is illustrated in Table 1. To ensure consistency across the integrated schema, we standardize units and data formats between the two datasets. All time-related measurements, such as latency and jitter, are represented in milliseconds (ms), while throughput values are expressed in megabits per second (Mbps). Latency values are stored as integers, whereas throughput, packet loss rate, and jitter are stored as floating-point numbers rounded to five decimal places. Timestamps are standardized to UTC with second-level precision.

Geo-location data is normalized using the ISO 3166-1 alpha-2 standard for country codes. For city and region names, we adopt the GeoNames *cities15000* dataset, which includes all cities worldwide with a population of 15,000 or more, alongside alternate names and relevant metadata such as administrative region and country identifiers. Regional names are sourced from the *admin1CodesASCII* dataset, which maps standardized region names to their corresponding countries. To ensure uniformity and avoid issues related to diacritics or non-Latin scripts, we exclusively use the ASCII versions of both city and region names [7].

To align Cloudflare AIM's *serverPoP* values with NDT7's city and country-based server location metadata, we utilize the publicly available airport dataset provided by *Datahub.io* [6]. This dataset allows us to map each IATA airport code to its corresponding municipality and country. Although this approach does not provide the exact physical location of the Cloudflare Radar server, we consider it a sufficiently accurate proxy for the purposes of estimating client-server proximity.

3.2.4 Cross-Source Considerations. While our unified schema enables side-by-side analysis of performance metrics from both Cloudflare AIM and MLab's NDT7 datasets, we recognize that these datasets differ significantly in measurement methodology, server infrastructure, and test execution. All metrics are collected under different network conditions and methodologies, making direct comparisons nontrivial. To ensure transparency and analytical integrity, we include a `data_source` field for every record, explicitly identifying whether a metric originates from Cloudflare AIM or NDT7. This enables source-aware analyses and helps avoid inappropriate cross-dataset aggregation.

Table 1: Proposed common schema for integrating NDT7 and Cloudflare AIM measurement data.

Field Name	Data Type
uuid	varchar(255)
test_time	UTC timestamp
client_city	varchar(255)
client_region	varchar(255)
client_country_code	character(2)
server_city	varchar(255)
server_country_code	varchar(255)
asn	integer
data_source	varchar(255)
packet_loss_rate	numeric(10,5)
download_throughput_mbps	numeric(10,5)
download_latency_ms	integer
download_jitter_ms	numeric(10,5)
upload_throughput_mbps	numeric(10,5)
upload_latency_ms	integer
upload_jitter_ms	numeric(10,5)

Additionally, to address the differences in network test environments and geographical routing behavior, we outline concrete preprocessing and normalization steps in Subsection 3.3 and discuss mitigating location-based biases in Subsection 3.4.

3.3 Data Preprocessing

To focus our analysis on the most relevant networks, we begin by filtering both datasets to retain only measurements from either (i) the **top five ISPs** in each country where Starlink operates, or (ii) **Starlink itself**, identified by autonomous system number *AS14953*. These top ISPs were identified using the query described in Subsection 3.1. By modifying the `SELECT` clause of that query, we generate a string formatted as an SQL `IN` clause, which is embedded into both the NDT7 and Cloudflare queries to restrict the data to the most relevant ASNs from the outset.

The *NDT7 dataset* records multiple metrics throughout the duration of each test. For consistency and completeness, we extract only the **final measurement**, which serves as a summary of the test and includes all relevant intermediate values. This filtering is implemented via a custom SQL query executed using the *Google BigQuery interface.*

To ensure data quality, we exclude incomplete or invalid entries. Specifically, we discard measurements where essential location metadata is missing (i.e., the client country code is null or empty), where throughput is recorded as zero, or where both download and upload measurements are null. The full query used for this extraction is provided in Appendix A, Listing 3.

In the *Cloudflare AIM dataset*, packet loss rate and jitter are reported as single values per test, whereas latency and throughput are measured multiple times during each test. Since these values are collected from the same client within a short time interval, they are typically similar. To avoid overrepresenting individual tests in the final analysis, we **aggregate** these multiple measurements into a single representative value.

We hypothesize that the **median** is the most suitable aggregation method, as it is resilient to outliers and better captures the central tendency in noisy measurements. The procedure for testing this hypothesis is detailed in Subsection 3.5. As with NDT7, we apply data quality filters to the Cloudflare AIM dataset. We discard entries with missing or empty client country codes to ensure that each measurement can be geolocated. The query used for this filtering and aggregation is provided in Appendix A, Listing 5.

To mitigate routing variability and reduce inconsistencies arising from distant or atypical server locations, we apply additional filtering based on client and server geo-location. The details of this filtering process are provided in Section 3.4.

3.4 Mitigating Network Location Bias

When a user initiates an Internet speed test, they are connected to a server capable of handling their request. The selection of this server is influenced by several factors, including: **(i)** load balancing to distribute traffic across multiple servers, **(ii)** network latency, aiming to minimize response time, **(iii)** geographic proximity to the client, **(iv)** server availability and health status to ensure reliability, and **(v)** routing policies determined by ISPs and the underlying network topology.

These factors can sometimes result in users being connected to geographically distant servers, especially during periods of high load or under specific routing conditions. This situation negatively impacts measured network metrics, particularly latency, since it is highly sensitive to the physical distance that data must travel. Such measurements introduce bias into the dataset, misrepresenting the user's typical experience.

To mitigate *geographical sampling bias*, we implement a server-based filtering strategy designed to retain only those measurements that reflect realistic client-server proximity. Using measurement data collected between March 1, 2025, and May 28, 2025—the most recent and representative period available—we identify, for each city, the set of servers that historically yielded the lowest observed latencies. This time-frame was chosen because it captures the latest state of Starlink's infrastructure, including major updates such as the activation of the Mozambique ground station in March 2025, which significantly impacted routing and latency patterns in neighboring countries like Zambia and Zimbabwe.

A measurement is retained only if it involves one of the "best" servers for the corresponding city. In cases where no historical measurements exist for a given city but are available at the country level, we fall back to using the country's best-performing servers. If no prior measurements are available for the entire country, we retain the measurement by default. While this fallback mechanism is not triggered in our current analysis—since the best-server selection is based on the same time window as the measurement data, ensuring all servers are within known countries—it may become necessary in a continuously deployed global telemetry system, where new regions might appear without recent baseline measurements and the best-server database must be regularly updated.

3.5 Distribution Comparison Procedure

To evaluate the feasibility of merging heterogeneous telemetry datasets, we develop a distribution comparison procedure between the NDT7 and Cloudflare AIM datasets. This approach aims to assess how similar their performance measurements are when appropriately preprocessed, particularly with respect to different aggregation techniques applied to Cloudflare AIM data.

Our analysis focuses on a one-week period, from **May 5 to May 11, 2025**. This timeframe represents typical Internet conditions across both weekdays and weekends, avoiding globally disruptive events such as international holidays or widespread outages. Although some localized events occurred during this period—such as public holidays in the UK and the Netherlands, and regional storms in Brazil and China—they are not deemed significant enough to distort the overall validity of the analysis.

Using the procedure described in Subsection 3.3, we create local PostgreSQL databases for both NDT7 and Cloudflare AIM. For mitigating geographical location bias, the server selection strategy described in Subsection 3.4 is used, which retains only measurements associated with the lowest-latency servers for each city.

Three aggregation techniques are considered for summarizing Cloudflare AIM measurements: the **mean**, the **median**, and the **90th percentile**. Each method is chosen for its distinct statistical properties: the mean captures the average performance but can be sensitive to outliers; the median provides a more robust central tendency measure, particularly in skewed or noisy datasets; the 90th percentile offers insight into near-worst-case performance, excluding the most extreme outliers.

To compare distributions between NDT7 and the aggregated Cloudflare AIM data, we select a subset of countries with a high number of both Starlink and terrestrial measurements across all datasets. After exporting these datasets as CSV files, we conduct our analysis using Python within a Jupyter Notebook environment.

For qualitative comparison, we plot the probability density functions (PDFs) of the selected metrics using histograms with consistent bin widths across datasets. These visualizations allow us to examine the general shapes of the distributions. For quantitative comparison, we compute the **Jensen–Shannon Divergence (JSD)**, a symmetric and smoothed measure of similarity between probability distributions, proven to be more accurate in measuring similarity between distributions than traditional information divergence [18]. The JSD is defined in Equation 1 based on the Kullback-Leibler divergence (Equation 2). Using a logarithm with base 2, the JSD is bounded between 0 and 1, where 0 corresponds to identical distributions and 1 indicates maximal divergence.

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2}\,\mathrm{KL}(P \parallel M) + \frac{1}{2}\,\mathrm{KL}(Q \parallel M), \tag{1}$$

where $M = \frac{1}{2}(P+Q)$ denotes the point-wise average of P and Q, and $\mathrm{KL}(\cdot \parallel \cdot)$ is the Kullback–Leibler divergence. The KL divergence (with logarithm base 2) is defined as:

$$\mathrm{KL}(X \parallel Y) = \sum_i X(i) \log_2 \frac{X(i)}{Y(i)}. \tag{2}$$

We use the implementation provided by the scikit-learn library [22], and consider JSD values below 0.1 as indicative of sufficient similarity to justify merging distributions. The Cloudflare aggregation method that results in the lowest average JSD across most countries is deemed the most representative for summarizing Cloudflare AIM data. The SQL queries used for generating each aggregation type are listed in Appendix A, and the full analysis—including data processing scripts and visualizations—is available in a public GitHub repository [29].

4 Integration

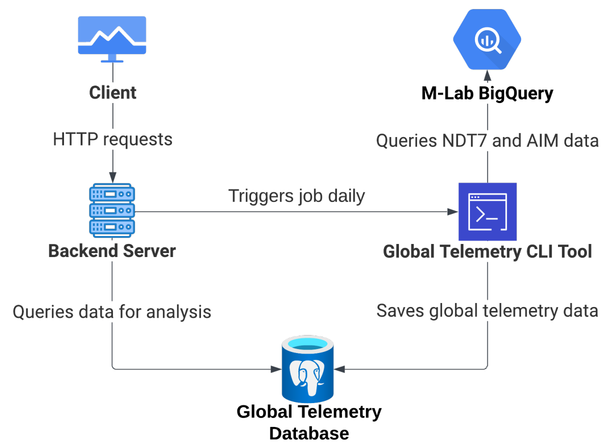


Figure 2: The integration between the Global Telemetry System (CLI tool) and the LEO Viewer (backend server)

The main system developed in this work is the **Global Telemetry System** [28], a command-line interface (CLI) tool designed to aggregate and standardize telemetry data from heterogeneous sources. This system is integrated into the **LEO Viewer** project [9], which provides a backend for comparing the performance of LEO satellite ISPs (such as Starlink) against terrestrial ISPs. The Global Telemetry System automates the data collection and preprocessing steps required for such comparisons.

As detailed in the *README.md* file of the CLI repository, the Global Telemetry System requires installation of the Google Cloud CLI and authentication with a Google account that has access to M-Lab's BigQuery dataset. While the dataset is public, specific steps are necessary to configure access. The tool also requires PostgreSQL database credentials, which are provided via a .env file. The CLI supports a variety of commands for data processing and maintenance:

- --init: Initializes the database schema and populates reference data.
- --date <date>: Processes telemetry data for a specific UTC date.
- --date-range <date_from>:<date_to>: Processes telemetry data over a specified UTC date range.
- --update-best-servers <date_from>[:<date_to>]: Updates mappings to the best servers.
- --update-countries-with-starlink <date_from>[:<date_to>]: Updates Starlink country coverage dataset.
- --update <CHOICES>: Refreshes reference datasets (ASNs, airports, cities).
- --drop: Drops all tables from the database.

Table 2: Median Cloudflare AIM (median-aggregated) network metrics before filtering, after filtering, and within the deleted records, for the week of 5-11 May 2025. Filtering is based on whether the measurement server was among the historically lowest-latency servers

Data	Packet Loss Rate	Download Throughput (Mbps)	Download Latency (ms)	Download Jitter (ms)	Upload Throughput (Mbps)	Upload Latency (ms)	Upload Jitter (ms)	Records
Before Filtering	0.00000	57.73731	40	22.27000	24.02869	47	20.84000	39,941
After Filtering	0.00000	60.18397	37	20.43000	26.10390	44	20.20000	30,207
Deleted	0.00000	51.96359	53	29.56000	18.02691	58	22.77000	9,734

Table 3: Median NDT7 network metrics before filtering, after filtering, and within the deleted records, for the week of 5-11 May 2025. Filtering is based on whether the measurement server was among the historically lowest-latency servers

Data	Packet Loss Rate	Download Throughput (Mbps)	Download Latency (ms)	Download Jitter (ms)	Upload Throughput (Mbps)	Upload Latency (ms)	Upload Jitter (ms)	Records
Before Filtering	0.00000	38.91341	67	3.01700	13.21189	75	12.86100	7,151,582
After Filtering	0.00000	38.14899	66	3.09100	12.70902	74	13.23700	6,696,934
Deleted	0.00000	50.44435	77	2.00600	20.08295	89	8.18500	454,648

Note: All date arguments must follow the `YYYY-MM-DD` *format.*

The most critical command is `--date`, which implements the methodology described in this paper. It retrieves all telemetry measurements from the top five ISPs in countries with Starlink data, along with Starlink's own measurements, for the specified UTC date. The data is standardized, filtered using server-based heuristics, and stored in the `unified_telemetry` table.

Figure 2 illustrates the system architecture. The backend server, part of the LEO Viewer project, triggers the `--date` job daily at 12:00 AM UTC. While Cloudflare continuously streams new measurements into the AIM dataset, NDT7 data for day *D* is only made available on day *D + 1*. Therefore, the 12:00 AM UTC trigger ensures that all records from the previous day have been ingested, making it the earliest safe point to avoid data loss.The processed telemetry data is then available for backend services that perform ISP performance comparisons.

5 Results and Analysis

5.1 Effect of Geographical Server Filtering

Tables 2 and 3 report the median values of various network measurements from the Cloudflare AIM (median-aggregated) and NDT7 datasets, respectively, before and after applying the server-based filtering technique described in Subsection 3.4. All data is from the week of 5–11 May 2025. In each table, the first row represents the measurements prior to filtering, the second row shows the results after filtering, and the third row provides statistics for the records that were excluded during the process.

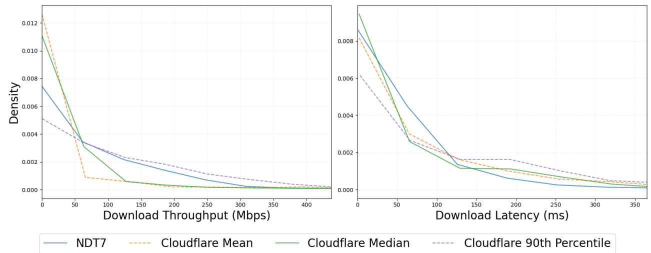
As shown in Table 2, the application of the filtering technique to the Cloudflare AIM dataset produces the expected outcome: the removal of measurements associated with atypical server assignments. The filtered-out data exhibits substantially higher latency—32.5% for download and 23.4% for upload—relative to the unfiltered dataset. When compared to the post-filtering data, the latency difference is even more pronounced, reaching 43.2% for download and 31.2% for upload. In addition to improved latency, jitter is reduced in both directions after filtering, while throughput shows a slight increase. Overall, the records excluded during filtering consistently exhibit inferior performance across all measured metrics, and their removal enhances the quality and reliability of the remaining dataset.

For the NDT7 dataset, the results summarized in Table 3 follow a similar pattern to those observed in the Cloudflare AIM dataset. The filtered-out data exhibits poorer network performance, with absolute increases in latency of approximately 10 ms for download and 14 ms for upload when compared to the unfiltered data. While these differences are comparable in magnitude to the Cloudflare results, the relative percentage changes are smaller due to slightly higher median latencies in the NDT7 data. Notably, although filtering removes measurements with very high latency, the improvement in overall performance is less pronounced, and throughput and jitter metrics slightly deteriorate after filtering.

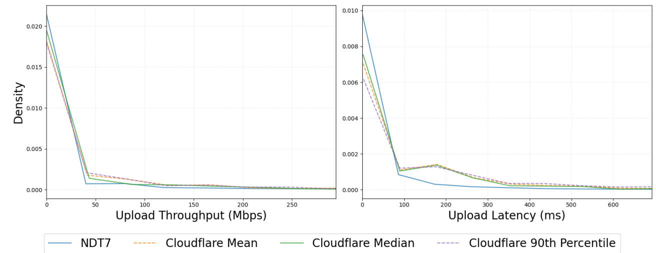
Approximately 24.4% of the records are discarded during filtering on Cloudflare AIM. This relatively large proportion can likely be attributed to the expansive global footprint of the Cloudflare Radar CDN infrastructure, which operates in over 330 cities across more than 125 countries [5]. The wide server distribution introduces higher variability in server assignment, thereby increasing the likelihood of noisy or inconsistent measurements. In contrast, only 6.4% of the records are removed from the NDT7 dataset. Given that M-Lab reports maintaining approximately 125 server locations [14], this observation supports the hypothesis that larger server deployments are associated with greater variability in server selection. However, further research is needed to confirm this relationship.

5.2 Distribution Comparison

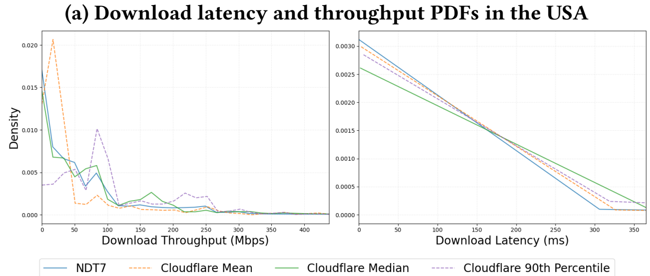
Figure 3 shows the probability density functions (PDFs) of upload and download latency and throughput, based on measurements collected during the week of 5–11 May, 2025. The analysis includes data from the top five terrestrial ISPs in each country with a Starlink presence, alongside Starlink itself. The distributions are derived from two sources: M-Lab's NDT7 dataset and Cloudflare AIM, with the latter aggregated using three methods: mean, median, and 90th



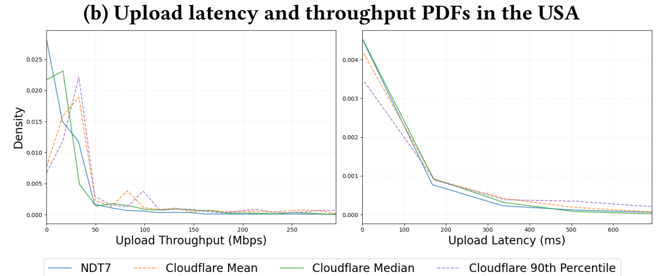
(a) Download latency and throughput PDFs in the USA



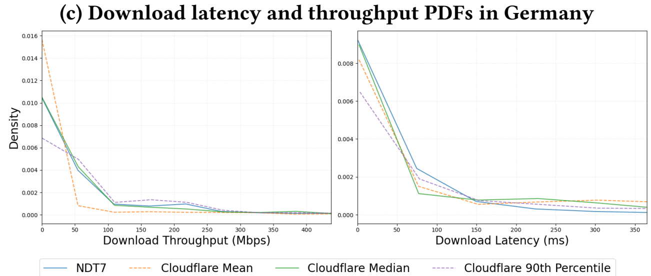
(b) Upload latency and throughput PDFs in the USA



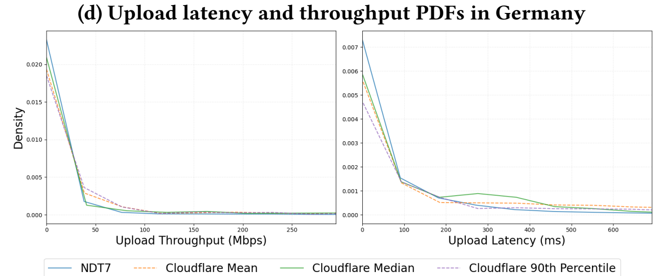
(c) Download latency and throughput PDFs in Germany



(d) Upload latency and throughput PDFs in Germany



(e) Download latency and throughput PDFs in Australia



(f) Upload latency and throughput PDFs in Australia

Figure 3: Probability density functions (PDFs) of upload and download latency and throughput measured during the week of May 5–11, 2025. Each row corresponds to a different country (USA, Germany, Australia), showing download (left) and upload (right) metrics. The plots compare M-Lab's NDT7 measurements with Cloudflare AIM data, aggregated using the mean, median, and 90th percentile.

percentile. Each row corresponds to a different country (USA, Germany, and Australia), selected for their geographic diversity (across three continents) and high measurement volumes from both terrestrial and satellite ISPs. Download metrics are displayed in the left column, while upload metrics appear in the right column. The x-axis is limited to the 95th percentile to highlight the region with the highest density of values.

The overall trend is consistent across all throughput plots (left-hand side of Figures 3a, 3b, 3c, 3d, 3e, and 3f). The Cloudflare AIM data aggregated using the 90th percentile exhibits a lower concentration of probability density in the lower throughput range (0–50 Mbps), deviating more noticeably from the NDT7 distribution. In contrast, both the mean and median aggregations align more closely with the NDT7 measurements, with the median providing the closest match overall. Notably, in Figure 3c, the median-aggregated Cloudflare data captures distinctive inflection points at approximately 45 Mbps and 80 Mbps, further indicating its ability to reflect the underlying NDT7 distribution. In Figures 3e and 3b, the Cloudflare AIM measurements aggregated using the median closely mirror the NDT7 distributions, demonstrating near-identical probability density functions.

For the latency plots (right-hand side of Figures 3a, 3b, 3c, 3d, 3e, and 3f), similar conclusions emerge. The 90th percentile aggregation exhibits lower density than NDT7 in the lower latency range (0–150 ms), but surpasses it in the higher ranges. The mean aggregation generally follows the median and NDT7 distributions closely; however, it tends to deviate slightly in the same direction as the median but to a greater extent. That is, when the median overrepresents or underrepresents NDT7, the mean typically amplifies that deviation. For example, it shows reduced density in the 0–200 ms range in Figure 3f, while in Figure 3d it shows slightly elevated density between 175 and 600 ms. Among the three, the median aggregation seems to align most closely with the NDT7 distribution, particularly evident in Figure 3c.

Quantitative results of the distribution comparison are presented in Table 4. The table reports the Jensen-Shannon Divergence (JSD) between the NDT7 samples and Cloudflare's data, aggregated using

Table 4: Jensen-Shannon Divergence (JSD) between NDT7 and Cloudflare datasets for download and upload latency/throughput, across selected countries, after server-based filtering, for the week of 5-11 May 2025

Country	Metric	Mean	Median	P90
USA	Download Latency	0.0171	0.0170	0.0565
	Download Throughput	0.0929	0.0520	0.0202
	Upload Latency	0.0459	0.0381	0.0720
	Upload Throughput	0.0191	0.0204	0.0206
Germany	Download Latency	0.0020	0.0058	0.0090
	Download Throughput	0.0671	0.0188	0.0917
	Upload Latency	0.0082	0.0087	0.0294
	Upload Throughput	0.0943	0.0334	0.1154
Australia	Download Latency	0.0322	0.0259	0.0584
	Download Throughput	0.0660	0.0086	0.0274
	Upload Latency	0.0349	0.0239	0.0646
	Upload Throughput	0.0251	0.0116	0.0318

three methods (mean, median, and 90$^{\text{th}}$ percentile) for both download and upload latency and throughput, across three countries. The median aggregation method yields the lowest JSD in 8 out of the 12 cases, and in three others, it performs nearly as well as the best method. These results support our hypothesis that the median is the most effective aggregation method. Notably, all aggregation strategies produce low JSD scores (all below 0.1), providing strong evidence that merging the two datasets is both reasonable and fair.

Table 5: Jensen-Shannon Divergence (JSD) between NDT7 and Cloudflare AIM datasets, across selected countries, after server-based filtering, for the week of 5-11 May 2025.

Country	Metric	JSD Value
USA	Download Jitter	0.1101
	Upload Jitter	0.0322
	Packet Loss Rate	0.0375
Germany	Download Jitter	0.0047
	Upload Jitter	0.0227
	Packet Loss Rate	0.0276
Australia	Download Jitter	0.0881
	Upload Jitter	0.0388
	Packet Loss Rate	0.0125

Table 5 presents the Jensen-Shannon Divergence (JSD) computed between the corresponding fields from the NDT7 and Cloudflare AIM datasets. It is important to note that jitter and packet loss metrics do not require aggregation and therefore remain consistent across the various Cloudflare datasets with different aggregation methods. The analysis is based on data from the week of 5-11 May 2025, following the application of the server filtering technique.

Overall, the results show generally low divergence values for jitter and packet loss, indicating strong agreement between the two datasets in capturing these key performance indicators. These findings are consistent with the low divergence values observed for latency and throughput metrics discussed in the preceding paragraphs, further validating the reliability of both measurement sources. Nonetheless, the relatively higher JSD observed for download jitter, particularly in the USA and Australia, emphasizes the complementary nature of the datasets and highlights the value of integrating multiple data sources to achieve a more comprehensive understanding of Internet performance. We provide insight into how Starlink compares against terrestrial ISPs in Appendix B.

6 Conclusions and Future Work

In this paper, we presented a methodology for harmonizing and integrating heterogeneous internet measurement datasets to enable fair and reproducible comparisons between LEO satellite ISPs, such as Starlink, and terrestrial alternatives. This methodology addresses schema inconsistencies, server-based geographic bias, and aggregation disparities by applying structured filtering, normalization, and statistical alignment techniques. We implemented this methodology in the form of the **Global Telemetry System**, a command-line tool that integrates data from M-Lab's NDT7 and Cloudflare's AIM datasets, which is used by the **LEO Viewer** backend for automated, daily data collection.

Our experiments, conducted using measurements from May 5–11, 2025, demonstrate that this approach enables consistent and meaningful comparisons between different ISP types. By combining structural unification with statistical consistency, our methodology provides a scalable foundation for future research and monitoring of next-generation internet infrastructure.

Limitations

Despite its contributions, our system some limitations. First, the pipeline is designed to run once per day, meaning that data is not processed in real-time—limiting its applicability for use cases that require live or near-real-time monitoring. Secondly, the data coverage is inherently skewed towards regions with greater internet test activity, which can limit the granularity and fairness of comparisons in underrepresented areas.

Future Work

There are several promising directions for future research. One aspect could involve extending the system to incorporate other public datasets, such as those from RIPE Atlas and Ookla, to improve robustness and geographic coverage. Additionally, further work could focus on identifying latent relationships within the telemetry data—for instance, investigating the impact of external factors such as weather conditions, time of day, or regional events on network performance. This would enable more nuanced filtering and debiasing strategies, enhancing the reliability of the dataset.

Another important direction is transforming the Global Telemetry System into a continuously running service with near real-time ingestion, visualization, and anomaly detection capabilities. A publicly accessible dashboard would significantly benefit researchers, policymakers, and network operators interested in tracking the evolving performance of LEO ISPs compared to terrestrial options.

Overall, this work represents a first step toward a global, reproducible, and extensible framework for assessing the quality of emerging satellite-based internet infrastructures.

7 Responsible Research

In accordance with the Dutch Code of Conduct for Research Integrity [10], we adhered to the core principles of honesty, scrupulousness, transparency, independence, and accountability. To ensure reproducibility, we (i) used only publicly available datasets (M-Lab's NDT7 and Cloudflare AIM via BigQuery, CAIDA AS-rank data, Datahub.io airports, GeoNames geopolitical data), (ii) fully documented every data-extraction and transformation step in our Jupyter notebooks, and (iii) published all code, database schema definitions, and SQL queries in our GitHub repository [29].

Both NDT7 and Cloudflare AIM datasets are made publicly available by M-Lab, with explicit informed consent obtained from users at test time. While the NDT7 dataset includes client IP addresses, users are clearly notified that their data, including IP information, will be published. This ensures ethical use under M-Lab's transparency policies,even though the data is not fully anonymized. In contrast, the CAIDA dataset used for ranking ISPs is fully anonymized, containing only top-level AS-flow aggregates collected on public backbone links, thus posing no identifiable privacy risk.

Throughout the writing of this paper and the associated coding work, generative AI tools, such as ChatGPT, were used solely to enhance the clarity, structure, and grammar of existing text. No code was generated or modified using AI, and no content was directly authored by generative models. All scientific reasoning, analysis, and implementation were conducted independently by the authors.

Acknowledgments

We thank Measurement Lab and Cloudflare Radar for providing us access to their datasets, through the Google BigQuery interface.

References

[1] Amazon. 2025. Project Kuiper: Satellite Internet Constellation. https://www.aboutamazon.com/what-we-do/devices-services/project-kuiper. Accessed: 2025-05-01.
[2] CAIDA. 2025. AS Rank Dataset Schema. https://catalog.caida.org/dataset/as_rank/schema. Accessed: 2025-05-03.
[3] CAIDA. 2025. CAIDA AS Rank GraphQL API Documentation. https://api.asrank.caida.org/v2/docs. Accessed: 2025-05-04.
[4] Inc. Cloudflare. 2025. Cloudflare AIM: Aggregated Internet Measurement. https://developers.cloudflare.com/speed/aim/. Accessed: 2025-05-01.
[5] Inc. Cloudflare. 2025. Cloudflare Global Network. https://www.cloudflare.com/network/. Accessed: 2025-06-02.
[6] DataHub. 2025. Airport Codes Dataset. https://datahub.io/core/airport-codes. Accessed: 2025-05-21.
[7] GeoNames. 2025. GeoNames Data Dump. https://download.geonames.org/export/dump/. Accessed: 2025-05-18.
[8] Vlad Graure. 2025. Top Five ASNs per Country Dataset. https://github.com/TUD-BScResearchProject-6079/global-telemetry-system/blob/main/Datasets/top-five-asns-per-country.csv. Accessed: 2025-06-18.
[9] Vlad Graure, Cristian Benghe, Janusz Urbanski, and Christiaan Baraya. 2025. LEO Viewer. https://github.com/TUD-BScResearchProject-6079/leo-viewer. Accessed: 2025-06-18.
[10] KNAW, NFU, NWO, TO2-federatie, Vereniging Hogescholen, and VSNU. 2018. Nederlandse gedragscode wetenschappelijke integriteit. https://doi.org/10.17026/dans-2cj-nvwu. Accessed: 2025-05-01.
[11] Lyntia. 2024. Telecommunications Technologies Helping Close the Digital Divide. https://www.lyntia.com/en/news/telecommunications-technologies-to-close-the-digital-divide/. Accessed: 2025-05-01.
[12] Measurement Lab (M-Lab). 2025. NDT7 (Network Diagnostic Test v7) Dataset. https://www.measurementlab.net/tests/ndt/. Accessed: 2025-05-01.
[13] Kyle MacMillan, Tarun Mangla, James Saxon, Nicole P. Marwell, and Nick Feamster. 2023. A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7). *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 7, 1 (2023), 19:1–19:26. doi:10.1145/3579448
[14] Measurement Lab. 2025. Developer Documentation. https://www.measurementlab.net/develop/. Accessed: 2025-06-08.
[15] Nitinder Mohan et al. 2024. A Multifaceted Look at Starlink Performance. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM. doi:10.1145/3589334.3645328
[16] RIPE NCC. 2025. RIPE Atlas: Global Internet Measurement. https://atlas.ripe.net/. Accessed: 2025-05-01.
[17] Ilana Newman. 2023. For Rural Communities, Broadband Expansion Is No Single Thing. https://www.governing.com/infrastructure/for-rural-communities-broadband-expansion-is-no-single-thing. Accessed: 2025-05-01.
[18] Fangyan Nie and Jianqi Li. 2022. Image Segmentation Method Based on Jensen-Shannon Divergence Thresholding. *IAENG International Journal of Computer Science* 49, 1 (2022), 1–7. https://www.iaeng.org/IJCS/issues_v49/issue_1/IJCS_49_1_21.pdf Accessed: 2025-05-21.
[19] OneWeb. 2025. OneWeb: Global Satellite Internet Constellation. https://www.oneweb.world. Accessed: 2025-05-01.
[20] Jianping Pan, Jinwei Zhao, and Lin Cai. 2023. Measuring a Low-Earth-Orbit Satellite Network. doi:10.48550/arXiv.2307.06863 arXiv:2307.06863 [cs.NI]
[21] SatelliteMap. 2025. Satellites Overview. https://satellitemap.space/satellites.html. Accessed: 2025-05-01.
[22] Scikit-Learn. 2025. Scikit-Learn Jensen-Shannon Divergence. https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.jensenshannon.html. Accessed June 18, 2025.
[23] SpaceX. 2025. Starlink Reaches 5 Million Subscribers. https://stories.starlink.com/. Accessed: 2025-05-01.
[24] SpaceX. 2025. Starlink Satellite Demisability. https://www.starlink.com/updates. Accessed: 2025-05-01.
[25] SpaceX. 2025. Starlink: Satellite Internet Constellation. https://www.starlink.com. Accessed: 2025-05-01.
[26] Hammas Bin Tanveer et al. 2023. Making Sense of Constellations: Methodologies for Understanding Starlink's Scheduling Algorithms. In *Proceedings of CoNEXT 2023*. ACM. doi:10.1145/3624354.3630586
[27] David Tuber. 2023. Measuring network quality to better understand the end-user experience. https://blog.cloudflare.com/aim-database-for-internet-quality/. Accessed: 2025-05-03.
[28] Vlad Graure. 2025. Global Telemetry CLI. https://github.com/TUD-BScResearchProject-6079/global-telemetry-cli. Accessed: 2025-06-18.
[29] Vlad Graure. 2025. Global Telemetry System – Scripts and Analysis. https://github.com/TUD-BScResearchProject-6079/global-telemetry-system. Accessed: 2025-06-18.
[30] Bingsen Wang et al. 2024. A Large-Scale IPv6-Based Measurement of the Starlink Network. *arXiv preprint arXiv:2412.18243* (2024). https://arxiv.org/abs/2412.18243

A SQL Queries

Listing 1: SQL Query for identifying the countries with Starlink Internet speed tests

```
SELECT Client.Geo.CountryCode AS country_code
FROM `ndt.ndt7`
WHERE date > '2023-01-01'
  AND client.Network.ASNumber = 14593

UNION DISTINCT

SELECT ClientCountry AS country_code
FROM `cloudflare.speedtest_speed1`
WHERE date > '2023-01-01'
  AND clientASN = 14593
```

Listing 2: SQL Query that indentifies the top five ISPs for each country with Starlink measurements

```
WITH rank_within_country_view AS (
        SELECT asn, RANK() OVER (PARTITION BY
            country_name ORDER BY rank ASC) AS
            rank_within_country
        FROM AS_Statistics
)
```

```
SELECT a.asn, a.asn_name, a.rank, b.
    rank_within_country, a.country_name, a.
    country_code
FROM as_statistics a
        JOIN rank_within_country_view b ON a.asn =
             b.asn
        JOIN countries_with_starlink_measurements
            c ON a.country_code = c.country_code
WHERE b.rank_within_country <= 5 OR a.asn = 14593
ORDER BY a.country_name, b.rank_within_country
```

Listing 3: SQL Query for extracting data from NDT7 dataset

```
SELECT
  a.UUID as uuid,
  DATETIME(TIMESTAMP(a.TestTime), "UTC") AS
      test_time,
  client.Geo.City AS client_city,
  client.Geo.Region AS client_region,
  client.Geo.CountryCode AS client_country_code,
  server.Geo.City AS server_city,
  server.Geo.CountryCode AS server_country_code,
  client.Network.ASNumber AS asn,
  ROUND(a.LossRate, 5) AS packet_loss_rate,
  CASE
    WHEN raw.Download.ServerMeasurements[
        SAFE_OFFSET(ARRAY_LENGTH(raw.Download.
        ServerMeasurements) - 1)].TCPInfo.RTT IS
        NOT NULL
    THEN ROUND(a.MeanThroughputMbps, 5)
    ELSE NULL
  END AS download_throughput_mbps,
  CAST(ROUND(raw.Download.ServerMeasurements[
      SAFE_OFFSET(ARRAY_LENGTH(raw.Download.
      ServerMeasurements) - 1)].TCPInfo.RTT /
      1000, 0) AS INT) AS download_latency_ms,
  ROUND(raw.Download.ServerMeasurements[
      SAFE_OFFSET(ARRAY_LENGTH(raw.Download.
      ServerMeasurements) - 1)].TCPInfo.RTTVar /
      1000, 5) AS download_jitter_ms,
  CASE
    WHEN raw.Upload.ServerMeasurements[SAFE_OFFSET
        (ARRAY_LENGTH(raw.Upload.
        ServerMeasurements) - 1)].TCPInfo.RTT IS
        NOT NULL
    THEN ROUND(a.MeanThroughputMbps, 5)
    ELSE NULL
  END AS upload_throughput_mbps,
  CAST(ROUND(raw.Upload.ServerMeasurements[
      SAFE_OFFSET(ARRAY_LENGTH(raw.Upload.
      ServerMeasurements) - 1)].TCPInfo.RTT /
      1000, 0) AS INT) AS upload_latency_ms,
  ROUND(raw.Upload.ServerMeasurements[SAFE_OFFSET(
      ARRAY_LENGTH(raw.Upload.ServerMeasurements)
      - 1)].TCPInfo.RTTVar / 1000, 5) AS
      upload_jitter_ms
FROM `measurement-lab.ndt.ndt7`
WHERE
  date >= '2025-05-05'AND date <= '2025-05-11'
  AND client.Geo.CountryCode IS NOT NULL AND
      client.Geo.CountryCode <> ''
  AND a.MeanThroughputMbps IS NOT NULL AND a.
      MeanThroughputMbps <> 0.0
  AND (
    raw.Download.ServerMeasurements[SAFE_OFFSET(
        ARRAY_LENGTH(raw.Download.
        ServerMeasurements) - 1)].TCPInfo.RTT IS
        NOT NULL
      OR
    raw.Upload.ServerMeasurements[SAFE_OFFSET(
        ARRAY_LENGTH(raw.Upload.ServerMeasurements
        ) - 1)].TCPInfo.RTT IS NOT NULL
  )
  AND client.Network.ASNumber IN (14593,
    --list continues with ASNs of top 5 ISPs from
        countries with Starlink measurements
    );
```

Listing 4: SQL Query for extracting data from Cloudflare AIM dataset, aggregating latency and throughput using the mean of all tests

```
SELECT
  measurementUUID AS uuid,
  DATETIME(TIMESTAMP(measurementTime), "UTC") AS
      test_time,
  clientCity AS client_city,
  clientRegion AS client_region,
  clientCountry AS client_country_code,
  serverPoP AS server_airport_code,
  clientASN AS asn,
  ROUND(packetLoss.lossRatio, 5) AS
      packet_loss_rate,

  ROUND((SELECT SUM(bps * bytes) / SUM(bytes) /
      1000000 FROM UNNEST(upload.bps) bps WITH
      OFFSET AS i JOIN UNNEST(download.bytes)
      bytes WITH OFFSET AS j ON i = j), 5) AS
      download_throughput_mbps,
  CAST(ROUND((SELECT AVG(lat) FROM UNNEST(
      loadedLatencyMs.download) lat), 0) AS INT)
      AS download_latency_ms,
  ROUND(loadedJitterMs.download, 5) AS
      download_jitter_ms,

  ROUND((SELECT SUM(bps * bytes) / SUM(bytes) /
      1000000 FROM UNNEST(upload.bps) bps WITH
      OFFSET AS i JOIN UNNEST(upload.bytes) bytes
      WITH OFFSET AS j ON i = j), 5) AS
      upload_throughput_mbps,
  CAST(ROUND((SELECT AVG(lat) FROM UNNEST(
      loadedLatencyMs.upload) lat), 0) AS INT) AS
      upload_latency_ms,
```

```
  ROUND(loadedJitterMs.upload, 5) AS
      upload_jitter_ms
FROM `measurement-lab.cloudflare.speedtest_speed1`
WHERE
  date >= '2025-05-05' AND date <= '2025-05-11'
  AND clientCountry IS NOT NULL AND clientCountry
      <> ''
  AND clientASN IN (14593,
    --list continues with ASNs of top 5 ISPs from
        countries with Starlink measurements
    );
```

Listing 5: SQL Query for extracting data from Cloudflare AIM dataset, aggregating latency and throughput using the median of all tests

```
SELECT
  measurementUUID AS uuid,
  DATETIME(TIMESTAMP(measurementTime), "UTC") AS
      test_time,
  clientCity AS client_city,
  clientRegion AS client_region,
  clientCountry AS client_country_code,
  serverPoP AS server_airport_code,
  clientASN AS asn,
  ROUND(packetLoss.lossRatio, 5) AS
      packet_loss_rate,

  ROUND((SELECT PERCENTILE_DISC(bps, 0.5) OVER ()
      FROM UNNEST(download.bps) AS bps LIMIT 1) /
      1000000, 5) AS download_throughput_mbps,
  CAST(ROUND((SELECT PERCENTILE_DISC(ltc, 0.5)
      OVER () FROM UNNEST(loadedLatencyMs.download
      ) AS ltc LIMIT 1), 0) AS INT) AS
      download_latency_ms,
  ROUND(loadedJitterMs.download, 5) AS
      download_jitter_ms,

  ROUND((SELECT PERCENTILE_DISC(bps, 0.5) OVER ()
      FROM UNNEST(upload.bps) AS bps LIMIT 1) /
      1000000, 5) AS upload_throughput_mbps,
  CAST(ROUND((SELECT PERCENTILE_DISC(ltc, 0.5)
      OVER () FROM UNNEST(loadedLatencyMs.upload)
      AS ltc LIMIT 1), 0) AS INT) AS
      upload_latency_ms,
  ROUND(loadedJitterMs.upload, 5) AS
      upload_jitter_ms
FROM `measurement-lab.cloudflare.speedtest_speed1`
WHERE
  date >= '2025-05-05' AND date <= '2025-05-11'
  AND clientCountry IS NOT NULL AND clientCountry
      <> ''
  AND clientASN IN (14593,
    --list continues with ASNs of top 5 ISPs from
        countries with Starlink measurements
    );
```

Listing 6: SQL Query for extracting data from Cloudflare AIM dataset, aggregating latency and throughput using the 90th percentile of all tests

```
SELECT
  measurementUUID AS uuid,
  DATETIME(TIMESTAMP(measurementTime), "UTC") AS
      test_time,
  clientCity AS client_city,
  clientRegion AS client_region,
  clientCountry AS client_country_code,
  serverPoP AS server_airport_code,
  clientASN AS asn,
  ROUND(packetLoss.lossRatio, 5) AS
      packet_loss_rate,

  ROUND((SELECT PERCENTILE_DISC(bps, 0.9) OVER ()
      FROM UNNEST(download.bps) AS bps LIMIT 1) /
      1000000, 5) AS download_throughput_mbps,
  CAST(ROUND((SELECT PERCENTILE_DISC(ltc, 0.9)
      OVER () FROM UNNEST(loadedLatencyMs.download
      ) AS ltc LIMIT 1), 0) AS INT) AS
      download_latency_ms,
  ROUND(loadedJitterMs.download, 5) AS
      download_jitter_ms,

  ROUND((SELECT PERCENTILE_DISC(bps, 0.9) OVER ()
      FROM UNNEST(upload.bps) AS bps LIMIT 1) /
      1000000, 5) AS upload_throughput_mbps,
  CAST(ROUND((SELECT PERCENTILE_DISC(ltc, 0.9)
      OVER () FROM UNNEST(loadedLatencyMs.upload)
      AS ltc LIMIT 1), 0) AS INT) AS
      upload_latency_ms,
  ROUND(loadedJitterMs.upload, 5) AS
      upload_jitter_ms
FROM `measurement-lab.cloudflare.speedtest_speed1`
WHERE date >= '2025-05-05'
  AND date <= '2025-05-11'
  AND clientCountry IS NOT NULL AND clientCountry
      <> ''
  AND clientASN IN (14593,
    --list continues with ASNs of top 5 ISPs from
        countries with Starlink measurements
    );
```

B Comparative Analysis of Starlink and Top ISPs in Selected Countries

B.1 Results and Interpretation

To highlight the comparative strength of our dataset in analyzing Starlink's performance relative to terrestrial ISPs, we present cumulative distribution function (CDF) plots for key network performance metrics. These metrics include download and upload latency, throughput, jitter, and packet loss rate, and were collected over the period of May 5–11, 2025. Our analysis focuses on a set of countries, namely the USA, Australia, Kenya, and Myanmar, selected for having substantial Starlink measurements.

For each country, we compare Starlink to the top five ISPs (where such data is available). Each figure contains one subfigure per country. Within each subfigure, the left-hand plot shows CDFs generated from Cloudflare AIM data, aggregated using the median, while the right-hand plot shows CDFs based on NDT7 data. The data sources and methodology used to process the data are detailed in Section 3.

In terms of latency, better performance corresponds to lower latency values, which manifest as faster growth in a CDF plot. We begin by examining download latency, illustrated in Figure 4. In the United States (Figure 4a), the NDT7 plot (right) suggests that Starlink experiences higher latencies than the top five ISPs. Conversely, the Cloudflare plot (left) shows that only two of ISPs outperform Starlink, while the others exhibit worse performance. In Australia (Figure 4b), Starlink generally performs worse than the leading ISPs, except for ASN-TELSTRA (AS1221), which Starlink outperforms in both NDT7 and Cloudflare measurements. A similar pattern is observed in Kenya (Figure 4c), where Starlink records higher latencies than the top ISPs, with the exception of CKL1-ASN (AS36926), which it outperforms in the Cloudflare AIM dataset. Finally, in Myanmar (Figure 4d), Starlink generally exhibits higher latencies than the top ISPs, with two exceptions: the Cloudflare data shows that at higher latencies (above 130 ms), Starlink outperforms TIMCL-AS-AP (AS136255), and the NDT7 data reveals a similar trend at latencies above 400 ms, where Starlink outperforms MPT-MM-AS-AP (AS45558).

Regarding upload latency, presented in Figure 5, the NDT7 data for the United States (Figure 5a, right) indicates that terrestrial ISPs generally achieve lower latencies; however, at higher latency values (around 60 ms), Starlink appears to perform better by concentrating more volume in that range and rising more rapidly toward one. A similar pattern is observed in the plots for Australia (Figure 5b) and Kenya (Figure 5c). In Myanmar (Figure 5d), the Cloudflare plot (left) suggests that Starlink struggles to achieve very low upload latencies but attains more medium latency values compared to terrestrial ISPs. Conversely, the NDT7 plot (right) reveals that Starlink is outperformed by terrestrial infrastructure in Myanmar.

In terms of throughput, slower growth in the CDF indicates a concentration of volume toward higher values, which corresponds to better performance. Regarding download throughput, highlighted in Figure 6, Starlink achieves average performance, its CDF being between the terrestrial ones. In Australia (Figure 6b), Starlink appears as the best performer; however, it should be noted that in the Cloudflare measurements, Starlink did not exceed 150 Mbps, while terrestrial ASes recorded values above 300 Mbps. In Kenya (Figure 6c), Starlink also appears to perform best, although its performance is quite similar to that of terrestrial ASes. Conversely, in Myanmar (Figure 6d), Starlink is clearly the worst-performing AS.

For upload throughput, depicted in Figure 7, Starlink consistently appears as the worst performer across all selected countries. Specifically, Starlink fails to exceed 42 Mbps in Australia, 25 Mbps in Kenya, and 60 Mbps in Myanmar, which is significantly lower than the over 100 Mbps achieved by most other ASes.

When considering jitter, better performance corresponds to lower values; hence, the CDF of a well-performing AS in terms of jitter should rise quickly towards one. We begin by examining download jitter, illustrated in Figure 8. In the United States (Figure 8a), Starlink outperforms NTT-DATA-2914 (AS2914) but generally exhibits higher jitter values than the other four ASes. An exception occurs in the Cloudflare AIM plot (left), where between 10 and 15 ms, Starlink's CDF surpasses those of other ASes except HURRICANE (AS6939), indicating that terrestrial providers achieve lower jitters overall, but Starlink performs better as jitter increases. In Australia (Figure 8b), Starlink demonstrates average performance, outperforming some ASes while being outperformed by others. Kenya (Figure 6c) and Myanmar (Figure 6d) depict Starlink as the worst performer, with occasional improvements at higher jitter values visible in the Cloudflare plots.

Figure 9 presents the CDFs for upload jitter across several countries. In the United States (Figure 9a), the NDT7 (right) data indicates that Starlink exhibits the worst performance, whereas in the Cloudflare AIM dataset (left), Starlink outperforms four of the five terrestrial ASes. In Australia (Figure 9b), the Cloudflare CDF shows Starlink as the best performer, while in NDT7, it attains average performance, with its CDF lying between those of the terrestrial providers. This pattern observed in Australia is similarly seen in Kenya (Figure 9c) and Myanmar (Figure 9d), where Starlink ranks as the best performer in the Cloudflare plots but achieves average performance in the NDT7 data.

For packet loss, a well-performing AS will typically exhibit lower loss rates. In terms of CDFs, this translates to curves that rise steeply and approach one quickly, indicating that most measurements fall at low loss rates. The CDFs for packet loss rates are shown in Figure 10. Across all four countries examined in this case study, we observe that Starlink's packet loss performance is comparable to that of the top terrestrial ASes. In both the Cloudflare and NDT7 datasets, Starlink consistently appears somewhere in the middle range of the plotted ASes, neither significantly outperforming nor underperforming its terrestrial counterparts.

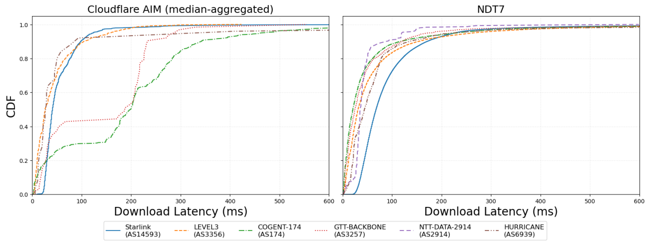
B.2 Conclusions and Limitations

Across the USA, Australia, Kenya, and Myanmar, Starlink shows mixed performance compared to top terrestrial ISPs:

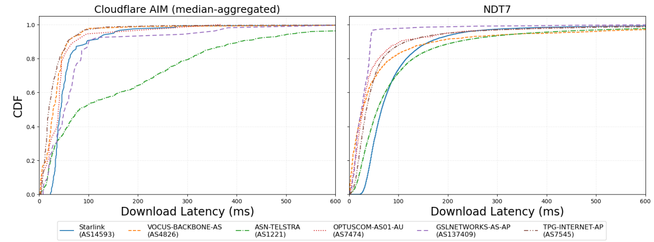
- **Latency:** Starlink generally lags behind terrestrial ISPs. However, it performs better at higher latency ranges in some countries.
- **Throughput:** Download throughput is average in USA, Australia and Kenya but weak in Myanmar. Upload throughput is consistently lower than terrestrial ISPs across all countries.
- **Jitter:** Results vary by dataset. Cloudflare data often shows Starlink as competitive, even best-performing, while NDT7 places it lower.
- **Packet Loss Rate:** Starlink performs on par with terrestrial ISPs, showing stable, mid-range behavior in all countries.

While Starlink rarely outperforms top ISPs, its jitter and packet loss results suggest it offers competitive service in areas with limited terrestrial infrastructure. The variation across datasets also underscores the value of multi-source measurements.

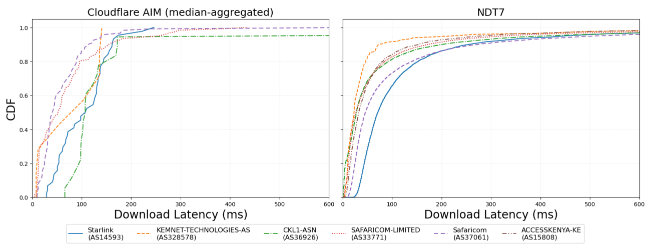
One important limitation of our case study stems from the relatively short data collection window and the lower popularity of Cloudflare's speed test compared to M-Lab's NDT7 test. As a result, the Cloudflare AIM dataset contains fewer measurements, leading to CDF plots that are often less smooth and, in some cases, not fully consistent with the NDT7-based plots.



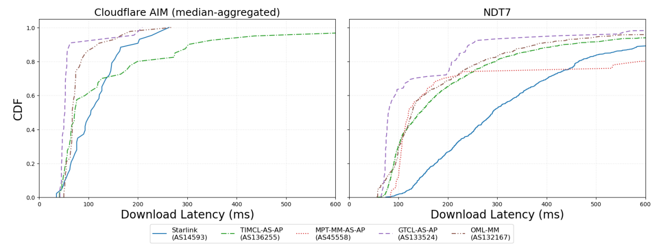
(a) USA



(b) Australia

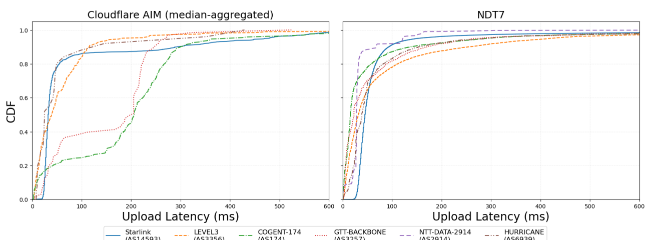


(c) Kenya

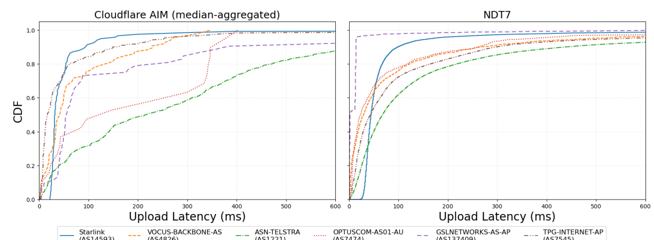


(d) Myanmar

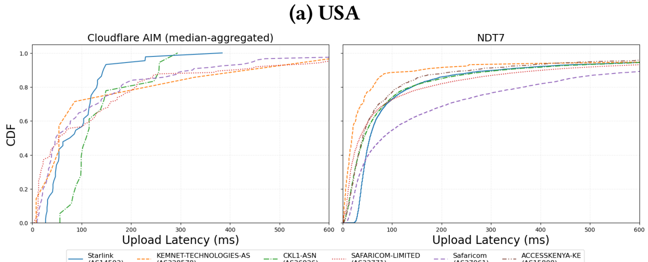
Figure 4: Cumulative Distribution Functions (CDFs) per country, comparing download latency for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.



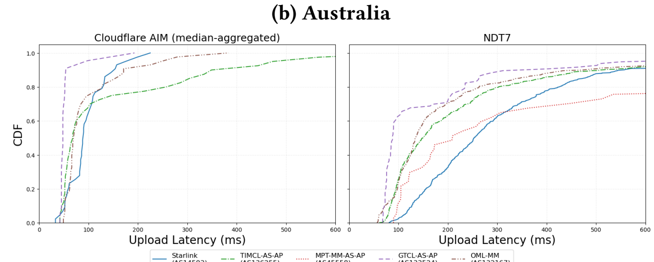
(a) USA



(b) Australia

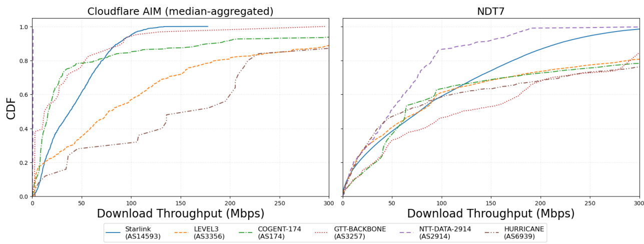


(c) Kenya

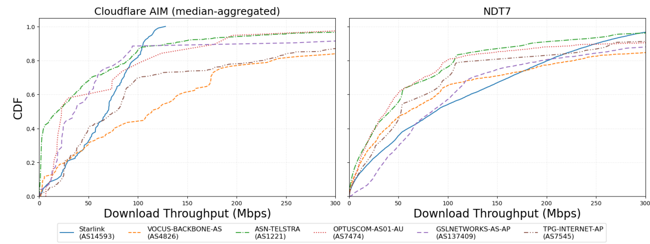


(d) Myanmar

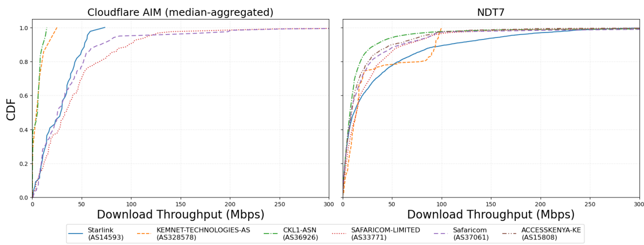
Figure 5: Cumulative Distribution Functions (CDFs) per country, comparing upload latency for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.



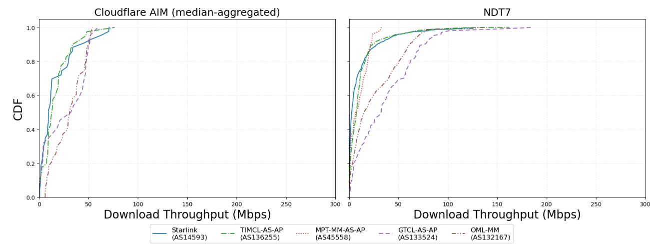
(a) USA



(b) Australia

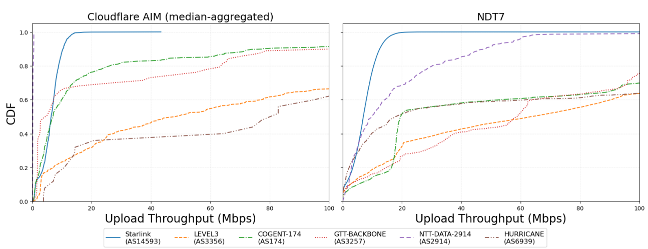


(c) Kenya

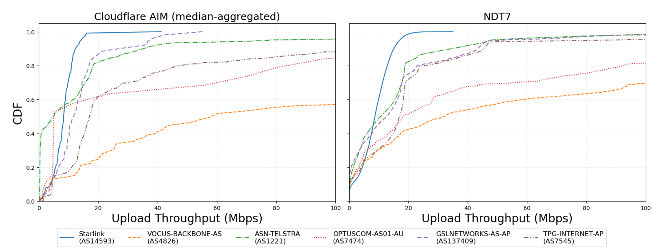


(d) Myanmar

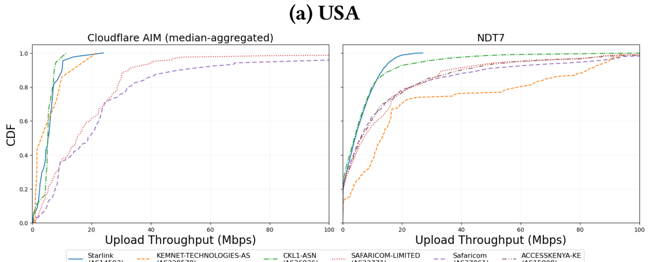
Figure 6: Cumulative Distribution Functions (CDFs) per country, comparing download throughput for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.



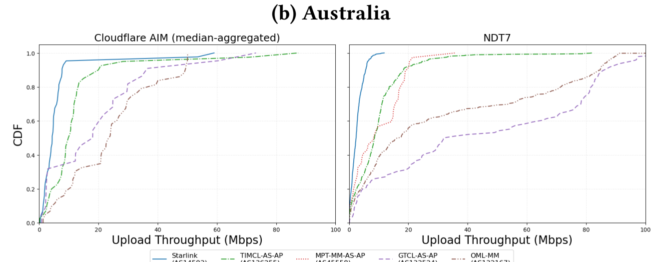
(a) USA



(b) Australia

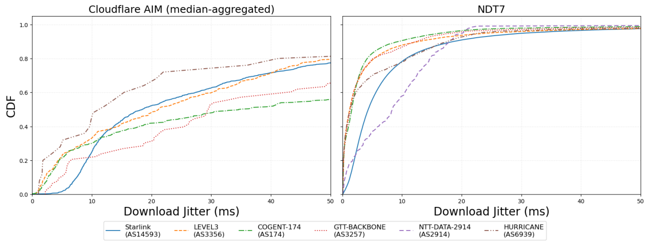


(c) Kenya

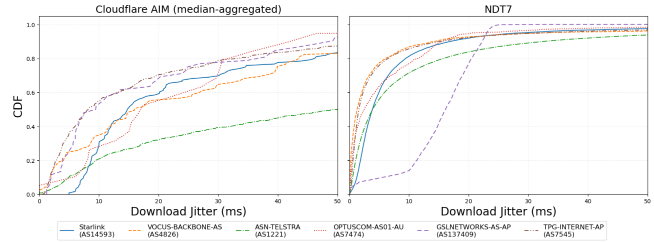


(d) Myanmar

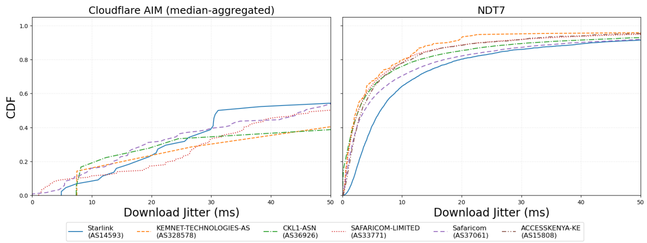
Figure 7: Cumulative Distribution Functions (CDFs) per country, comparing upload throughput for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.



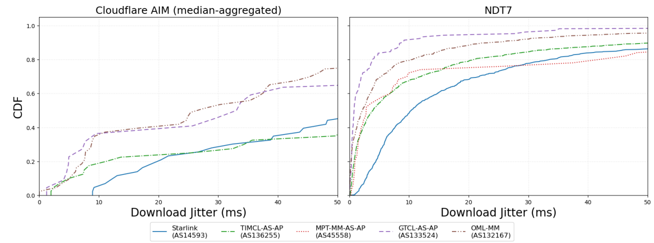
(a) USA



(b) Australia

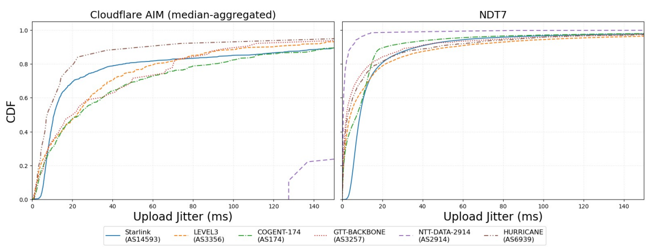


(c) Kenya

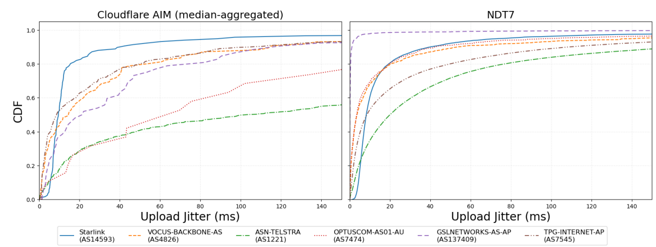


(d) Myanmar

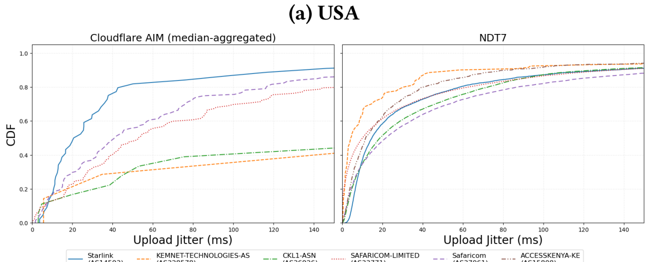
Figure 8: Cumulative Distribution Functions (CDFs) per country, comparing download jitter for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.



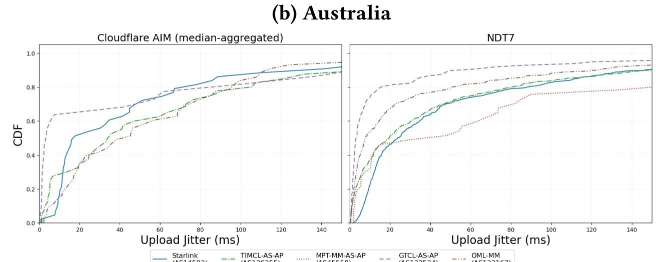
(a) USA



(b) Australia

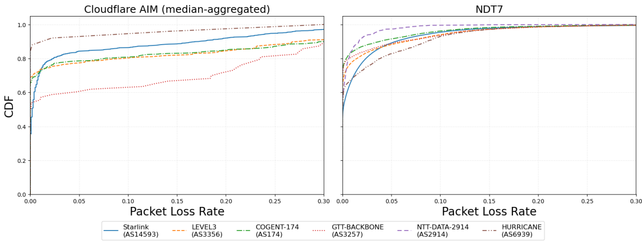


(c) Kenya

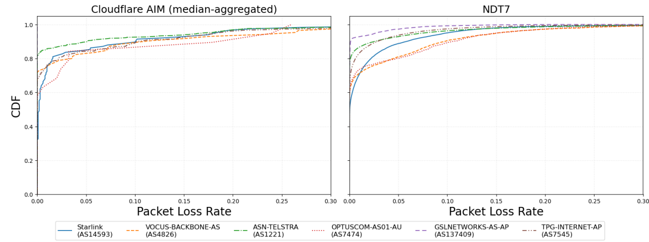


(d) Myanmar

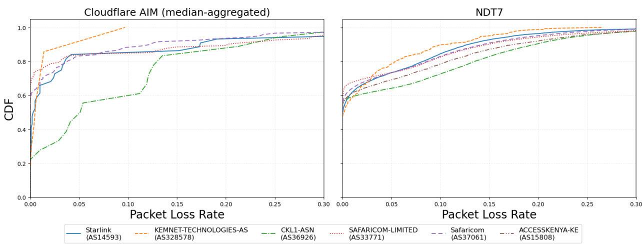
Figure 9: Cumulative Distribution Functions (CDFs) per country, comparing upload jitter for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.



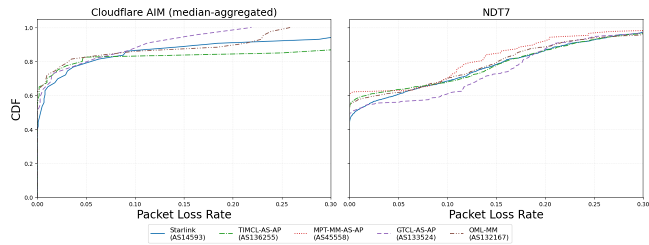
(a) USA



(b) Australia



(c) Kenya



(d) Myanmar

Figure 10: Cumulative Distribution Functions (CDFs) per country, comparing packet loss rate for Starlink (ASN 14593) and the top five ASNs. Each subfigure represents a different country, with the left plot showing Cloudflare AIM data (aggregated using the median) and the right plot showing NDT7 data. Data collected during the week of 5–11 May 2025, after server-based filtering.