When satellite is all you have: watching the internet from 550 ms

When Satellite is All You Have: Watching the Internet from 550 ms

Daniel Perdices
Universidad Autónoma de Madrid
daniel.perdices@uam.es

Gianluca Perna
Politecnico di Torino
gianluca.perna@polito.it

Martino Trevisan
University of Trieste
martino.trevisan@dia.units.it

Danilo Giordano
Politecnico di Torino
danilo.giordano@polito.it

Marco Mellia
Politecnico di Torino
marco.mellia@polito.it

ABSTRACT

Satellite Communication (SatCom) offers internet connectivity where traditional infrastructures are too expensive to deploy. When using satellites in a geostationary orbit, the distance from Earth forces a round trip time higher than 550 ms. Coupled with the limited and shared capacity of the physical link, this poses a challenge to the traditional internet access quality we are used to.

In this paper, we present the first passive characterization of the traffic carried by an operational SatCom network. With this unique vantage point, we observe the performance of the SatCom technology, as well as the usage habits of subscribers in different countries in Europe and Africa. We highlight the implications of such technology on Internet usage and functioning, and we pinpoint technical challenges due to the CDN and DNS resolution issues, while discussing possible optimizations that the ISP could implement to improve the service offered to SatCom subscribers.

KEYWORDS

Satellite Communications, Performance, Passive Measurements

ACM Reference Format:

Daniel Perdices, Gianluca Perna, Martino Trevisan, Danilo Giordano, and Marco Mellia. 2022. When Satellite is All You Have: Watching the Internet from 550 ms. In *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22), October 25–27, 2022, Nice, France.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

While 5G and Fiber-To-The-Home (FTTH) technologies give us access capacity on the order of Gb/s [15, 30, 34, 43] and Content Delivery Networks (CDN) can guarantee end-to-end delay of less than a millisecond [1, 20, 40], there are significant parts of the world where economic and technological constraints force people to rely on solutions that provide far more constrained access to the Internet. These include mountainous and rural areas in developed countries, as well as the entire territory of underdeveloped countries, where even the supply of stable electricity can be problematic. In such scenarios, Satellite Communications (SatCom) offers a practical connectivity solution. Among the available SatCom technologies, geostationary (GEO) satellites are the oldest and most widely used solution [17], with the first offerings dating back to the early 2000. Here, a satellite orbits the Earth at an altitude of about 36 000 km, and moves at an angular velocity equal to the Earth's rotational speed. To an observer on Earth, the satellite appears immobile, making it easier to establish a communication link. A single GEO satellite can cover entire continents, and the directional beams enable efficient space and frequency multiplexing, with each beam providing a total channel capacity on the order of 10 Gb/s [21].

IMC '22, October 25–27, 2022, Nice, France
2022. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

In addition to limited shared capacity, GEO satellite communications suffer from high propagation latency, which is about 550 ms for a round trip (including two passes through the satellite). For this reason, complicated Medium Access Control (MAC) and scheduling protocols coordinate the access and sharing of the satellite's uplink and downlink, while traffic shapers, Performance Enhancing Proxies (PEP), and TCP optimization solutions attempt to mitigate the effects of end-to-end delay and limited capacity [4, 32, 44].

In this paper, we have the unique opportunity to present the first large-scale passive characterization of a global GEO SatCom Internet access solution. Through passive instrumentation of the satellite ground station, we observe traffic from tens of thousands of customers in more than 20 countries in Europe and Africa. On the one hand, this allows us to characterize the different Internet usage habits in the different scenarios, if any. On the other hand, we observe the impact of SatCom technology on performance and identify possible optimization strategies.

The main observations can be summarized as follows:

- In Africa, chat and social media applications consume 100 and 10 times more data than in Europe. This is due to the presence of community WiFi points that share SatCom access.
- Since these applications are accessed throughout the day, the typical peak time in African countries is anticipated in the morning.
- Satellite channel protocols and solutions increase the Round Trip Time (RTT) much more than just the propagation delay. Link channel quality and congestion (if any) can actually add seconds to the end-to-end RTT.
- In SatCom networks, all traffic must pass through the same ground station, in our case in Europe. This impacts local popular services. For example, services in Africa suffer from the additional delay caused by traffic being routed back and forth through the ground station.
- To complicate the picture, most customers use open DNS resolvers, some of which are located in China and Africa. This increases DNS response time to hundreds of milliseconds

and jeopardizes the server selection policies of CDNs and DNS resolvers.

We believe that the characterization offered in this paper contributes to understanding the complexity of the Internet by providing a novel perspective on SatCom networks and customers. We also discuss possible technical improvements that SatCom providers might consider to improve the quality of service offered to their customers. These include the use of additional ground stations to route traffic more efficiently and the control of DNS requests or responses to limit the impact of incorrect server selection. To allow the research community to conduct experiments with an emulated GEO SatCom connection and compare it to other connection technologies (including the novel Starlink connection with data from [26]), we have created a data-driven model for our ERRANT network emulator tool [41] and make it available at https://github.com/SmartData-Polito/errant.

In the remainder of the paper, we introduce the SatCom technology and monitoring infrastructure (Section 2) and provide an overview of our dataset (Section 3). We then illustrate our results in terms of volumetric traffic distribution (Section 4), user habits and service consumption (Section 5), and performance indicators (Section 6). Finally, we discuss related work (Section 7) and draw conclusions (Section 8).

2 MEASUREMENT SETUP AND METHODOLOGY

In this section, we describe our measurement setup and the methodology we use to gather and analyze data from the actual deployment of a large international SatCom operator. We first provide an overview of the specific data-link technologies that can affect Internet access performance, but avoid detailing the complexity of the physical layer of SatCom transmissions. In this section, we describe our measurement setup and the methodology we use to gather and analyze data from the actual deployment of a large international SatCom operator. We first provide an overview of the specific data-link technologies that can affect Internet access performance, but avoid detailing the complexity of the physical layer of SatCom transmissions.

2.1 The SatCom Network

As in traditional SatCom networks, the operator has deployed a satellite infrastructure consisting of satellites in geostationary orbit, and a ground infrastructure. Referring to Figure 1, subscribers employ a dedicated equipment, i.e., the customer-premises equipment (CPE), to connect their devices (PC, smartphone, etc.) to the SatCom network. The CPE consists of a dish antenna and a router/modem that manages the satellite links and access protocols on the one side, while offering WiFi and Ethernet connectivity on the other side.[1] The satellite acts as relay for subscriber traffic, which traverses 35 786 km twice to reach the ground station, accumulating from 240 ms to 280 ms, depending on the location on Earth of the subscriber. The ground station terminates the satellite segment and forwards the traffic to the Internet.[2] Notice that this forces all traffic to enter the internet from the location where the ground station is. In our measurement setup, we monitor the traffic managed by one satellite in geostationary orbit. This satellite offers service in Europe and Africa, from Ireland to South Africa. At the time of data collection, the satellite operator operates a single ground station in Italy, through which all traffic passes to reach all Internet services.

The satellite is equipped with multiple directional antennas, each managing a transmission beam that points to a specific region on Earth. This allows the reuse of frequencies to increase overall capacity while optimizing the use of spectrum reserved for satellite communications. Each beam acts as a separate and independent physical channel, providing aggregate capacity on the order of Gb/s, the actual capacity being configurable. Two separate beams (and frequencies) cover each area, one for the uplink (from users' CPE to the satellite) and one for the downlink (from the satellite to users' CPE). A separate beam pair also connects the satellite and the ground station.

On the shared uplink channel, the transmission of packets involves a complicated MAC protocol: a slotted-Aloha protocol allows the CPE to access the shared reservation channel the first time it needs to transmit. Then, a Time Division Multiple Access (TDMA) scheduling protocol run by the satellite allocates time-slots to each active CPE to avoid collisions and to fairly share capacity among the active users at each TDMA frame. The satellite then forwards the packets to the ground station via a dedicated high-capacity beam.

On the downlink channel, the ground station transmits the packets directly to the satellite, which then forwards them to the destination CPE by selecting the correct frequency and beam. In this case, the packets are broadcasted to all receivers, which filters those destined for their CPE MAC address and discard the packets destined for other CPEs. In addition to the TDMA and MAC schemes, Forward Error Correction (FEC) and Automatic Repeat Request (ARQ) mechanisms provide a reliable data-link service. All in all, these proprietary algorithms provide a reliable, almost error-free, bi-directional point-to-point link between each CPE and the ground station. By combining these MAC, scheduling, FEC and ARQ protocols, further random delays are added to the communication between the CPE and the ground station.

To mitigate the potential performance degradation caused by this high latency, the SatCom operator relies heavily on a Performance Enhancing Proxy (PEP) to improve TCP performance on the satellite segment. A PEP is a network component that improves end-to-end performance by transparently manipulating TCP connections. Defined in RFC 3135 [16], PEP works as follows in our case. In the lower part of Figure 1, the subscriber CPE acts as a transparent TCP proxy for the end user's TCP traffic. It terminates all TCP connections initiated by applications on end-user devices and forwards TCP payload to the ground station via a bidirectional reliable tunnel over UDP. In detail: When a subscriber's device initiates a new TCP connection via a SYN packet, the CPE impersonates the destination server and immediately completes the TCP

[1] A customer may represent a single individual, a household, a company's office, or a community-based WiFi internet access solution.

[2] The total round trip time (RTT) of any communication is in the order of 550 ms since the packets must go through the satellite link on both the forward and backward paths.

three-way handshake, allowing the client application to send the initial data with no delay.

Here, the CPE acts as a L4 proxy. It buffers the TCP data stream and forwards it to the ground station via the bidirectional UDP tunnel at the allowed uplink rate. The ground station again works as a L4 proxy. When it receives a *Connect* request from the CPE, it establishes a new TCP connection to the actual destination server. It then forwards the data to/from the CPE via the downlink/uplink satellite tunnel. In this way, the TCP congestion control algorithm is effectively decoupled, allowing the ground station proxy to retrieve the data from the origin server at the backbone-path rate and the CPE to forward the data to the end user's device as quickly as possible. Note that the download rate of the ground station from the origin server is still regulated by the download rate of the end device because the buffer capacity of PEP in the ground station is limited. Note that user traffic using UDP (e.g., DNS, QUIC) cannot benefit from PEP acceleration and therefore UDP packets are forwarded as is.[3]

At last, the ground station also acts as a Network Address Translation (NAT) box, DNS resolver, and supports Quality of Service (QoS) schedulers to prioritize and shape traffic depending on the application. To this end, the SatCom operator uses L3/L4 and domain name-specific rules to prioritize interactive traffic and shape video streaming flows. The shaper allows also to enforce commercial maximum capacity of up to 5 Mb/s in the uplink, and 10, 20, 30, 100 Mb/s in the downlink based on the subscriber's contract.

In our setup, the SatCom operator provides private IPv4 addresses to each customer CPE. This means that all connections must be initiated by an end-user client and no server can be run on the customer's premises.[4]

2.2 Passive Measurements

We instrument the SatCom operator's network to collect passive measurements of all subscriber traffic. To this end, we deploy a passive probe at the operator's ground stations in Italy. Here we collect all traffic after the operation of the PEP, which handles traffic in the satellite segment. We observe all packets exchanged by each customer – that we uniquely identify by their SatCom CPE IPv4 IP address. Using a router span port, we mirror both downlink and uplink traffic to a high-end measurement server equipped with two Intel X710 network cards. The server runs Tstat[39], a custom flow monitoring software that generates rich per-flow summaries in real time from the processed data packets. To handle the high rate, our software resorts to the Data Plane Development Kit (DPDK) library for packet capture [18], which enables accelerated packet processing by bypassing the kernel-space drivers and protocol stack, and guarantees that all packets are processed in real time without information loss. Using the classic 5-tuple, Tstat identifies and tracks the evolution of TCP and UDP flows. For each flow, it extracts hundreds of statistics for both flow directions. The metrics we mainly rely on are: the i) flow size and duration, ii) the timing information of the first 10 packets, iii) the server and client IP address, iv) the TCP RTT between data and ACK segments, and v) the name of the contacted server as retrieved by the SNI, HTTP or DNS protocol..

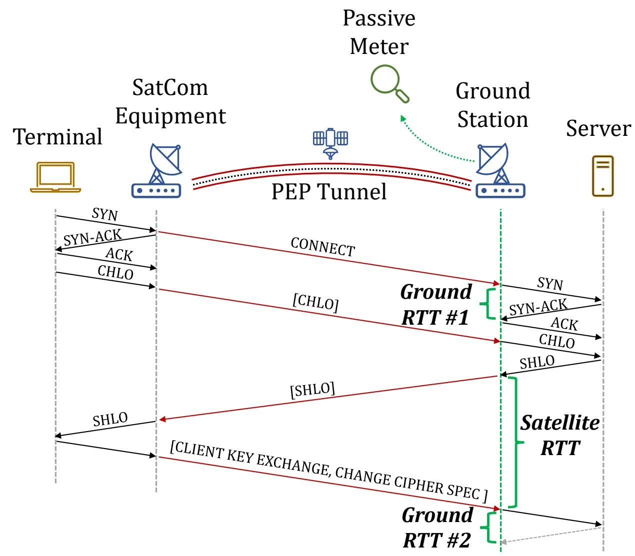


Figure 1: Methodology for the estimation of the Satellite Segment RTT.

The measurement of the TCP RTT deserves a careful explanation because the presence of PEP makes the measurement of RTT particularly troublesome. Indeed, the PEP causes the total RTT to be divided into three components, as shown in the bottom part of Figure 1: (i) *home RTT* – between the user device and the user's SatCom CPE; (ii) *Satellite RTT* – between the CPE and the ground station, where TCP segments are forwarded over the satellite PEP tunnel; and (iii) *ground RTT* – between the PEP terminator at the ground station and the destination server. Our vantage point is co-located with the ground station and therefore observes traffic exchanged from the PEP proxy to the Internet. For the ground RTT measurement (iii), TSTAT uses the TCP connection initiated by the ground station PEP. For each TCP segment sent, it measures the time to the corresponding ACK, calculating the minimum, maximum, average, and standard deviation of all RTT samples in a TCP flow. To measure the RTT (ii) of the Satellite segment, we need an additional ingenuity. Specifically, we leverage the TLS handshake of a TCP data flow to measure the time from the Server Hello message to the next Client Key Exchange message/Change Cipher Spec message. This time also includes the Home RTT (i), which we can consider negligible compared to the satellite. In this way, and only for TLS flows completing the TLS negotiation, we can safely estimate the delay caused by the satellite at least once per flow.

At last, for both TCP and UDP, the software runs a Deep Packet Inspection (DPI) module that identifies the most popular protocols and extracts various information from headers. In particular, it annotates each flow with the server *domain name* as extracted from the *Host* header in case of plain-text HTTP, or from the *Server Name Indication* (SNI) field in the case of TLS or QUIC flows.[5] For DNS traffic, the software logs each requested domains and obtained

[3]The PEP can only act as a L4 proxy without violating the authenticity provided by TLS.
[4]The SatCom provider offers hosting of servers in the data center for customers interested in running services.
[5]We use the term *domain* meaning Fully Qualified Domain Name.

responses, including the DNS server IP address the client used to resolve the name.

2.3 Ethical Aspects

Passive monitoring involves capturing and processing traffic generated by human beings, thus we need to take proper actions to protect as much as possible the individual's privacy. Indeed, IP address is considered a Personally Identifiable Information (PII) and it can be used to identify and track individuals. The characteristics of traffic, such as the list of visited websites, can be considered Sensitive Personal Information (SPI), as they can reveal personal aspects and habits of an individual.

For this work, we take all possible countermeasures to properly handle our measurements. First, the setup, management and data collection were physically managed uniquely by the operator personnel, who control the data collection process. Second, we configured the data collection to limit as much as possible the exposed information. We process packets in real time and save only strictly required information in flow logs. In details, we do not store any information present in headers that can be associated with a single user. Customers IP addresses are anonymized in real time using the CryptoPan algorithm [10] which preserve the subnet structure of the original IP addresses. The only sensitive information remains the server IP addresses and the domain names customers' visited, which we process only to extract aggregated statistics for the most popular services. Third, we only have access to the anonymized logs that we store in our secure Hadoop cluster, which is not reachable from the public internet and has strictly controlled physical access. The operating system and software are kept up to date to avoid possible vulnerabilities, and strict user access policies limit the access to the data only to authorized users.

The operator SatCom Data Protection Officer (DPO) has approved the above process and we have verified with our institutional review board that the data we collect is exempt from their approval.

3 DATASET PROCESSING AND OVERVIEW

In this paper, we consider all data collected during February-April 2022, resulting in 4.3 PB, for a total of 34.4 billion flows. On a daily basis, we transfer the flow summaries from the measurement server located in the SatCom provider premises to the Hadoop storage cluster, where we post-process the data using Apache Spark with custom designed analytics to compute various statistics and distributions.

3.1 Data enrichment and aggregation

In the first step of processing, we enrich the data by adding information about the customer's country (obtained by mapping the encrypted customer subnet to the corresponding country with the support of the SatCom operator) and about the service offered by the server. We focus here on six classes of services: Video Streaming, Social Networks, Audio Streaming, Chat, Work-related applications, and Search Engines. We rely on custom regular expressions that map popular server names to services. In detail, for each service class, we enumerate the top and local players by manually inspecting the list of most popular domains by volume and popularity. For each service, we enumerate the list of fully-qualified domains and second-level domains used to serve its content.[6] In some cases, we use regular expressions to generalize the set of domains. This is the typical case of CDN server names, which often include numbers or country codes in the domain. We report the full list of domains and regular expressions in the Appendix A. We uniquely use the domain to classify the service, as we do not capture the full HTTP URL, which is typically encrypted within the TLS session. For TLS flows, we obtain the domain from the SNI field of the Client Hello messages. As a result, we are sometimes unable to distinguish between sub-services from the same provider (e.g., Google Search and Maps both share the SNI `*.google.com`). As an alternative to manually curated lists, we could rely on online ranking and analytics tools (e.g., Alexa, Cisco Umbrella, or Similar Web). However, these services are known to list only the main domain of a given service, while not listing the domains of third party services such as content-delivery providers and support services used by the first-party service. Given the small number of services we are interested in, we choose to create the lists manually.

Table 1: TCP/UDP traffic breakdown by protocols.

Protocol	Volume share
TCP/HTTPS	56.0 %
TCP/HTTP	12.1 %
Other TCP	7.0 %
UDP/QUIC	19.6 %
UDP/RTP	1.1 %
UDP/DNS	$<$ 0.1 %
Other UDP	4.2 %

The second step is to create aggregated views of the data to obtain traffic breakdowns by protocols, server domains, time (with 1hour granularity), country of the customer, and contacted service. This aggregation step facilitates subsequent data processing by reducing the amount of data to be processed by several orders of magnitude, enabling real-time data exploration.

3.2 Dataset Overview

We first present an initial overview of the dataset by presenting the breakdown by country and protocol. Table 1 summarizes the latter. As expected, web traffic accounts for most of the traffic, with HTTPS and QUIC accounting for 56 % and 19.6 % of the total volume, respectively. HTTP still accounts for 12.1 % share. This is consistent with other studies that showed the convergence of Internet protocols towards encrypted web protocols [12, 35, 40]. Interestingly, despite the high latency due to the satellite link, we observe a non-negligible amount of video or voice traffic using Real Time Protocol (RTP).

Looking at traffic by country, we observe a large imbalance in the number of customers and thus traffic, as shown in Figure 2. The blue bars indicate the share of data traffic per country. The red line shows the share of customers. The countries are sorted by decreasing total volume. Interestingly, the two figures are not completely proportional and show that customers in African countries

[6] We handle the case of two-label top level domains – e.g., `co.uk`.

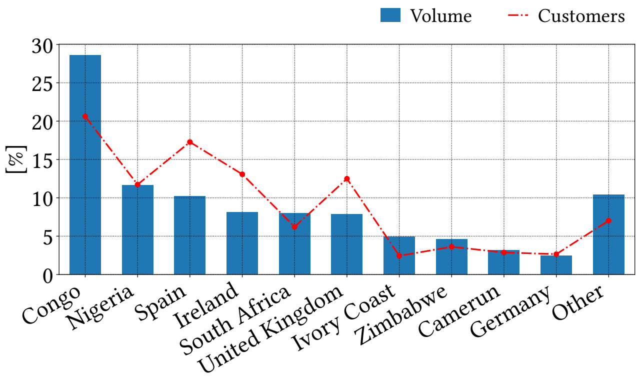


Figure 2: Per country breakdown of traffic volume and user base.

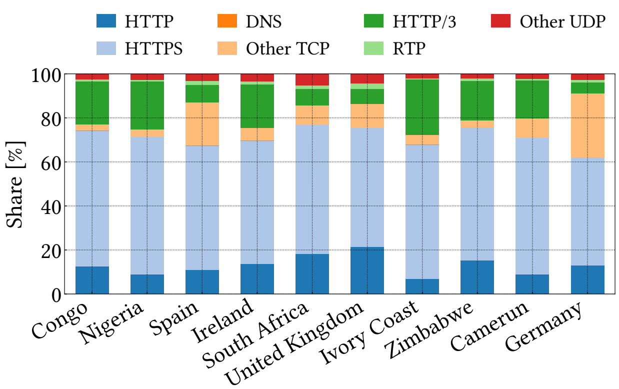


Figure 3: Protocol share per country.

consume much more traffic on average than customers in European countries. For example, Congolese customers are 20 % of overall customers, but they generate 27 % of volume (each generates about 600 MB per day). Spaniards are about 16 % of customers, but generate only 10 % of volume (each generates only 170 MB per day). This suggests that African customers may share Internet connections with multiple end-users, while European customers may resort to SatCom access only when forced to do so. In section 5 we will explore this direction in more detail.

To complete this overview, we list the breakdown of protocols by country. We limit the analysis to the top-10 countries. Figure 3 shows the results that exhibit some considerable differences. In European countries, a large portion of TCP traffic is not due to web protocols. The case of Germany is extreme: 35 % of all TCP traffic is due to other protocols. Manual inspection suggests that this is due to the use of Virtual Private Network (VPN) solutions (unknown protocol, non-standard port, long-lived flows without parallelism).

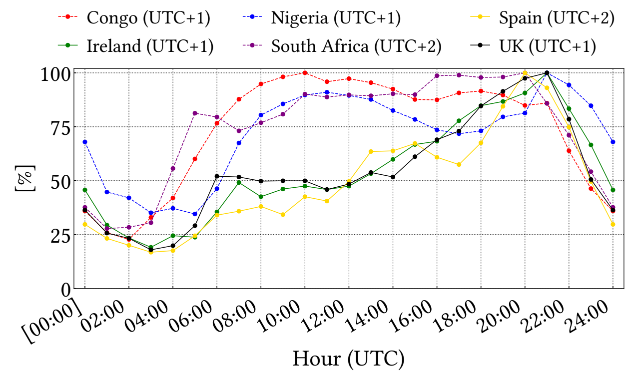


Figure 4: Daily trends per country.

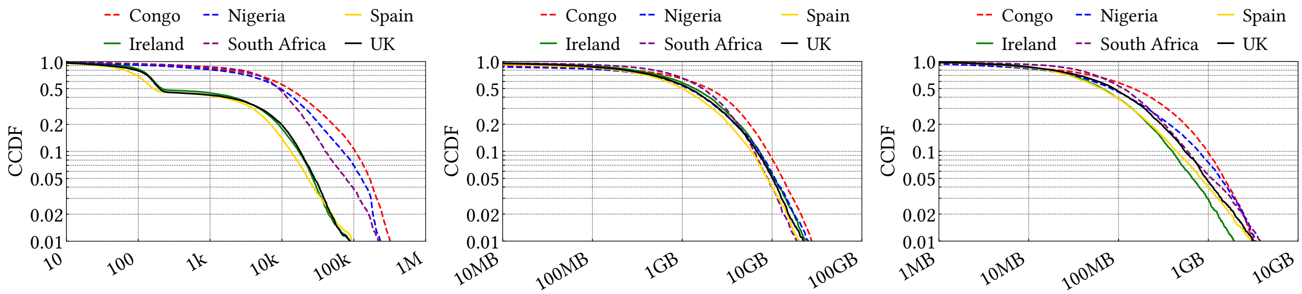
Note also that the percentage of unencrypted HTTP traffic is higher in Ireland and the U.K. than in other countries. This is due to popular Microsoft and Sky services that use HTTP to distribute software updates and video content. Conversely, Congo, Nigeria, and South Africa show very similar protocol breakdown. This may reflect the different customer base in Europe and Africa, with some business customers in the former (confirmed by the SatCom operator) using non-web-based protocols for VPN and internal services.

Given this initial overview, we limit our analysis below to the top 3 countries in Europe and the top 3 countries in Africa to compare usage, performance and services used.

4 HOW MUCH CUSTOMERS CONSUME

In this section, we discuss the temporal pattern and volume of traffic generated by customers in different countries. We start our analysis from the traditional hourly traffic pattern, which we report in Figure 4. The y-axis reports the percentage of traffic volume at a specific hour, normalized over the maximum value for the given country. For each time bin, we report the average value seen at that time during the whole time period, summing the upload and download traffic. We use the UTC time zone and countries in different time zones appear shifted.

We immediately observe how African (dashed lines) and European (solid lines) countries exhibit very different traffic patterns. In Europe, the traffic peak happens during evening prime time between 18:00-20:00-UTC. Conversely, during the day the traffic volume settles to lower values, down by 50 % in the morning and as low as 20 % at night. Conversely, in African countries, we observe a much higher traffic consumption during the morning too. For Congo (dashed red line), the absolute peak is at 9:00-UTC (10:00 local time). In Nigeria and South Africa, the morning peak reaches 90 % of evening peak time. Notice also at night the low-peak that is almost as high as 40 % of peak-time. This suggests that customers' use of the SatCom access differs between Europe and Africa, hinting a classic leisure usage for the former.



(a) Flows (all customers)

(b) Download volume (active customers)

(c) Upload volume (active customers)

Figure 5: Distribution of daily volume per customer in different countries. Notice the log scale on both axes.

We now focus on the volume of traffic generated per each customer, per each day. In Figure 5, we report the empirical Complementary Cumulative Distribution Function (CCDF), using log-log scales. We consider both the total number of flows and the total bytewise volume generated in each day by a given customer. We consider both TCP and UDP, and distinguish upload and download volume. Note that a single subscription generates many samples in the distribution, one per day.

Start from the number of daily flows per customer shown in Figure 5a, and focus on the main body of the CCDF (the leftmost part of the figure). We clearly observe that a significant fraction of European subscribers (solid lines) generate less than some hundred flows in 24 hours. In fact, the curves show a clear knee between 50 and 250 flows – i.e., more than 50 % of customers generate less than 250 flows in a given day. Those flows are likely being generated by the SatCom equipment or devices left connected to the network but unused. This could be the case of customers buying satellite access for their second houses in the remote regions that they use only during holidays. Conversely, in African countries (dashed lines) this effect is not present. From now on, we define *active customers* those that generate at least 250 flows per day.

Focus now on the tail of the curve. Here we can clearly see that African customers generate almost an order of magnitude more flows per day than European customers. This is due to the presence of some WiFi access points that share SatCom access in community internet solutions or internet cafés. In fact, most Africans in Congo and more in Nigeria and South Africa than in Europe, lack home Internet access. Therefore, people have to go to public places – Internet cafes, libraries, workplaces, etc. – to access the Internet. This situation was observed by the scientific community as early as 2013 [37] and more recently in 2019 [28]. The multiplexing of several end-users behind a single customer CPE IP address results in an inflated number of per-customer daily flows.

Next, we characterize the amount of daily downloaded and uploaded amount of traffic by each active customer. We show results in Figure 5b and 5c, respectively, for downlink and uplink. First, we observe that African customers download more data than European customers. Yet, the increase is reduced. For instance, in Congo, the percentage of heavy hitters (those customers downloading more than 10 GB in a day) is twice as much (8 %) than in Spain (4 %). This different behavior clearly impacts the congestion on the SatCom link, and the per-customer *cost* the SatCom operator in terms of volume.

	Congo	Nigeria	South Africa	Ireland	Spain	UK	Average
Google	62.96	61.26	64.72	68.58	68.30	65.48	64.20
Whatsapp	61.22	51.18	62.88	59.59	63.82	53.75	58.62
Snapchat	33.93	28.90	19.14	38.52	12.33	28.50	28.10
Wechat	6.42	3.55	1.11	0.49	0.06	0.41	2.99
Telegram	1.83	3.17	1.28	0.53	1.75	0.29	1.64
Instagram	48.81	41.04	40.67	48.53	45.59	40.43	44.84
Tiktok	41.56	31.99	36.31	40.11	31.89	36.53	36.95
Netflix	17.34	17.84	38.91	50.91	39.20	46.41	30.21
Primevideo	3.90	3.77	8.42	21.30	22.78	28.21	11.94
Sky	15.71	7.86	7.26	27.68	6.04	28.37	14.87
Spotify	37.78	30.31	33.19	46.79	45.20	39.73	38.15
Dropbox	11.50	9.22	16.57	10.39	9.34	16.81	11.75

Figure 6: Heatmap of the service popularity in different countries.

Interestingly, the difference between European and African countries is more pronounced in upload volume than in download volume. A Look at Figure 5c shows that Congo, Nigeria and South Africa have 10 %, 7 % and 5 % heavy hitters (those customers uploading more than 1 GB of data in a day), respectively, compared to less than 3 – 4 % in the U.K., Spain and Ireland. As we will see in the next session, customers who tend to upload a lot of content tend to exhibit a large usage of instant messaging applications, likely sharing images and videos from their mobile app. Overall, we find that SatCom customers generate similar traffic volumes as FTTH and ADSL customers. Compared to some recent work [40], the average download (upload) volume per FTTH customer was on the order of 1 GB (100 MB) per day in 2017. We thus observe a significant increase in the volume of traffic exchanged by customers despite the limited possibilities offered by SatCom access.

5 WHAT CUSTOMERS CONSUME

We now examine the habits of SatCom subscribers in terms of services they access. As described in Section 2, we identify the services by examining the domain of TLS, HTTP, and QUIC flows.

For each country, we extract the list of popular services, manually create regular expressions to identify them from the domain, and assign a category among Audio steaming, Chat, Search engine, Social, Video streaming, and Work.

In Figure 6, we first give, for each category, the percentage of customers accessing different services on a daily basis, separated by country. We focus on a subset of the services for which we can write regular expressions that match domains that we know the user intentionally visited. For example, we do not report on Social Media Networks or widely used services (e.g., YouTube) because these services often appear as third-party services in web pages (e.g., embedded social buttons, videos, tracking services).

Among Chat services, WhatsApp is the most widely used service - comparable to Google, which is the most popular service, as expected. This is consistent with our earlier findings [40]. Interestingly, Snapchat ranks second, while in Congo more than 6 % of customers also use WeChat for communication. This suggests the presence of Chinese-related communities. Telegram has yet to gain momentum.

Instagram and TikTok have similar penetration in all countries, with the latter being only 4 – 7 percentage points less popular than Instagram.

In the paid-video category, it is notable that they are more popular in Europe than in Africa, with only South Africa achieving similar penetration. This is likely due to both economical and cultural effects, as well as the investments these platforms make in each country. For completeness, as previously said, Sky uses HTTP rather than HTTPS for serving the video content, and its popularity in Ireland and the U.K. leads to the increase in HTTP traffic observed in Figure 3.

Our findings corroborate recent research on African Internet usage, which shows the strong presence of social and chat services. For example, a recent report from the Pew Research center [36] shows that chat and social networking are much more popular than paid services in Africa. In fact, paid video streaming services are not yet very popular in Africa, with Netflix having an estimated 2 million users on the continent, according to the 2022 annual report from Digital TV consulting [33]. However, the same report predicts rapid growth, which is also confirmed by a report from Conviva, a major video distribution company [7]. We note that South Africa is peculiar among African countries and that the strong penetration of streaming in South Africa is already well known.[7]

Next, we focus on the volume of traffic generated when accessing the different service categories per country. Here we consider all services, assuming that social buttons, tracking cookies, etc., consume little volume compared to customers using the actual service. Figure 7 uses boxplots to show the distribution of daily volume per customer accessing each category. The box extends from the lowest to the highest quartile, with a line at the median. Whiskers that extend from the box show the 5^{th} and 95^{th} percentiles.

Audio streaming services consume the least amount of traffic in Africa and slightly more in Europe, where some customers consume more than 50 MB in some days. Chat application usage is much more heterogeneous. While customers in Europe consume a median of less than 10 MB per day for Chat services, this value

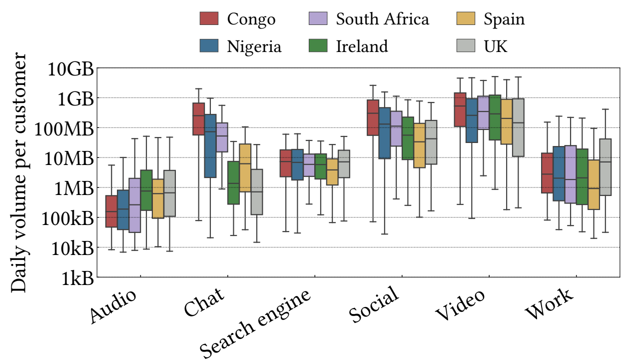


Figure 7: Boxplot of the daily volume consumption per customer when accessing different service category.

surprisingly increases by more than three orders of magnitude in African countries. Customers in Congo have a daily median of 250 MB, with the top-5 % of the heaviest customers consuming more than 2 GB on some days. These are likely those community WiFi Access Point (AP) that share SatCom access with multiple end users. The same effect is observed in the Social Media category, with a daily median of 300 MB in Congo, but only 30 MB in European countries. In contrast, the differences are smaller in the Video streaming category. However, the share of video traffic comes from different services: free YouTube in Africa, and paid video streaming services in Europe.

6 WHICH PERFORMANCE CONSUMERS GET

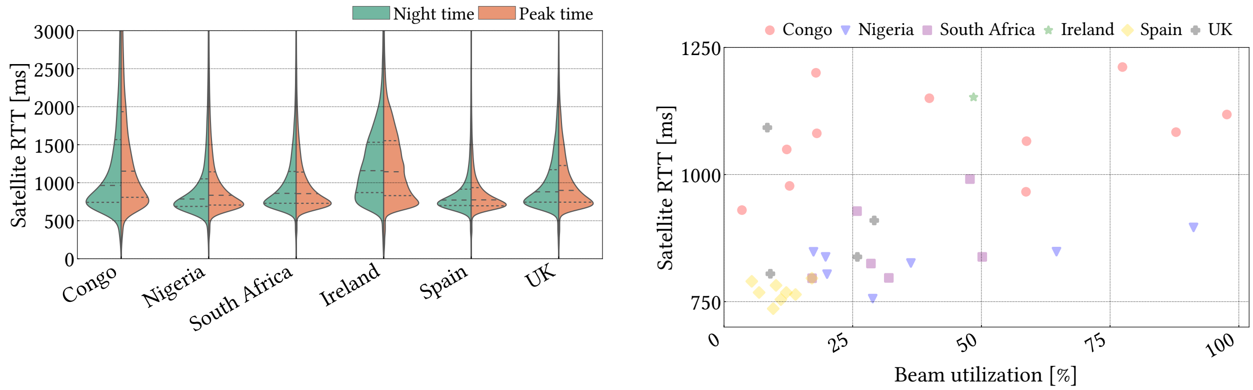
We now discuss the performance SatCom customers get. We focus on classical Quality of Service (QoS) indicators, namely RTT and throughput, and finally drill down on DNS performance.

6.1 Satellite RTT analysis

Here, we examine how the satellite access link and SatCom network architecture affect end-to-end RTT. We consider the satellite RTT and the ground RTT separately, as we defined in Section 2.

Focus first on the satellite RTT shown in Figure 8a. We report the measurements at night – when we expect low congestion on the satellite link – and at peak time. The dashed lines indicate the median, while dotted lines are used for the 25th and 75th percentiles. As expected, the minimum satellite RTT is above 550 ms. However, the distributions show very large variability with RTT that can be higher than 2 s and varies widely in each country. This variability is due to several factors: Queuing delay at various forwarding elements; Processing and transmission delay for limited performance terminals; Packet losses and TCP retransmission; but the main reason is the SatCom access technology. Specifically, Spain has the best RTT at night in general, with 82 % of RTT samples less than 1 s. Nigeria follows in second place and has even better RTT than Spain at low values. This is due to Nigeria favorable position, where the satellite is closer to the zenith (and thus has a shorter line of sight

[7]https://www.finder.com/za/streaming-statistics



(a) Distributions and quartiles of the satellite RTT per country for night time (from 2:00 to 5:00 local time) and peak time (from 13:00 to 20:00 local time).

(b) Median satellite RTT per beam.

Figure 8: Satellite RTT computed from TLS handshake.

than Spain). South Africa and U.K. suffer from a larger zenith angle and thus a larger RTT.

In contrast, Congo and Ireland suffer from a much higher and more variable RTT. For Congo, the main cause is the congestion on the satellite beams covering the country. In such bandwidth-constrained scenarios, the MAC protocol and the PEP scheduler may delay the transmission of packets by several frames, affecting the satellite RTT. For example, note that about 20 % of RTT samples are longer than 2 s. The high RTT values already occur during periods of low peak traffic and worsen during periods of high traffic, with RTT values increasing significantly - compare median values. For completeness, note that congestion also affects some Nigerian beams, while it is practically unnoticeable in Spain, U.K. and South Africa.

For Ireland, on the other hand, the different shape of the CDF and the practically identical RTT during nighttime and peak hours rule out congestion as the main cause of RTT variability. In fact, Ireland is located at the edge of the satellite coverage area with a large zenith angle, so the satellite transmission channel suffers from severe transmission impairments. The Satellite data-link protocol must deal with such impairments, which affect access time for those customers in particularly unfavorable locations.

To give more details, Figure 8b shows the median satellite RTT for each beam and relates it to the beam utilization[8]. We consider the peak time interval and mark each beam with the corresponding country it serves. Both Congo and Ireland suffer from high delay almost regardless of beam utilization. Nigeria, Spain and U.K. exhibit in general lower per-beam RTT. When checking this with the SatCom provider, the staff confirmed that there is some congestion that occurs, but it is not due to the beam capacity, but rather to the saturation of the PEP processing ability. This, in turn, slows down the forwarding of packets, especially during the initial phase of the connection setup. The amount of PEP resources that the

[8]We normalize the results to the maximum utilization observed across all beams to avoid disclosing the actual per-beam utilization.

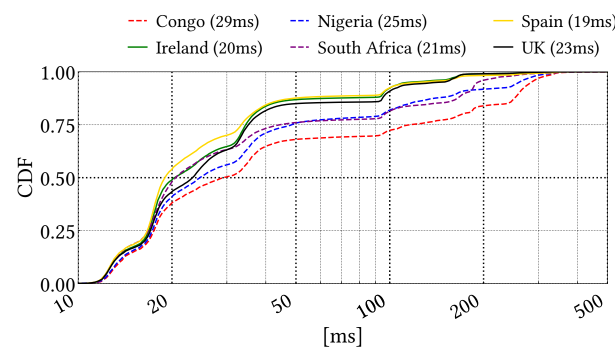


Figure 9: Ground segment RTT computed as the average RTT in each TCP flow. Legend details the median.

SatCom provider allocates to each beam and country depends on the SLA and the cost of the service. This clearly shows the overall complexity of the SatCom access technology, with implications on end-user quality of experience.

6.2 Ground RTT analysis

We now focus on the ground RTT shown in Figure 9. This RTT considers only the part of the path between the SatCom ground station and the server in the Internet.

In general, the ground RTT is much more deterministic than the satellite RTT. Here, clear bumps reflect the proximity of the servers on the Internet to the SatCom ground station. Focus on the European countries. The closest group has an RTT of about 12 ms and serves about 20 % of the traffic. These are CDN nodes of well-known players with widespread infrastructures with which the SatCom provider has direct peering agreements. A second group of servers is around 15 − 17 ms, and a third around 35 ms. All of these servers are located in Europe and serve more than 80 % of

	Congo	Nigeria	South Africa	Ireland	Spain	UK	Median Response Time [ms]
Operator-EU	0.87	9.10	1.87	43.75	28.95	38.10	3.98
Google	85.68	50.69	63.47	38.49	61.27	34.67	21.98
CloudFlare	3.02	2.54	10.36	2.03	2.05	6.04	19.97
Nigerian	0.00	11.84	6.32	0.00	0.00	0.00	119.98
Open DNS	1.22	4.00	0.65	0.49	0.72	6.97	17.99
Level3	0.45	7.63	0.09	0.00	0.00	0.49	23.99
Baidu	0.68	0.32	0.22	0.12	0.11	0.05	355.97
114DNS	2.97	3.43	1.64	0.05	0.03	0.01	109.98
Other	5.11	10.46	15.38	15.07	6.87	13.67	29.97

Figure 10: Adoption and median response time of DNS resolvers.

	UK		Nigeria		
	Op-EU	**Google**	**Op-EU**	**Google**	**114DNS**
captive.apple.com	19.1 ms	26.0 ms	23.1 ms	38.4 ms	110.4 ms
play.googleapis.com	16.3 ms	17.7 ms	38.7 ms	36.0 ms	114.2 ms
*.nflxvideo.net	-	25.5 ms	33.6 ms	28.8 ms	20.1 ms

Table 2: Average ground segment RTT per country and DNS resolver.

the traffic for European customers. Continuing this analysis, we observe another group of servers at about 95 ms (180 ms). Most of these servers are cloud servers located in the East (West) coast of the U.S., which suffer from the latency of crossing the Atlantic Ocean (and the U.S.). This reflects the typical RTT on the common Internet paths [1, 20, 40].

Now look at the RTT for African countries. Surprisingly, they have a higher RTT than their European counterparts. Since all traffic must be routed through the same ground station in Europe, African countries experience additional ground RTT when the final server is located in the original African country, e.g., when the end-user accesses a local service that is not served by global CDNs. In other words, the location of the ground station in Italy forces all traffic to be routed through Italy. This creates the rightmost bumps, where RTT on the ground increases to 300 − 400 ms.

By manually examining the services offered by these servers, we can confirm that they are likely popular services in the country of origin. Again, we also observe a significant proportion of Chinese services that are particularly popular in Congo (note the last bump in the ground RTT). This is related to the presence of Chinese companies in the country.

The SatCom provider is well aware of the RTT inflation due to the forced routing through the single ground station in Italy. They are already evaluating the possibility of setting up a ground station in Africa to optimize traffic routing and reduce groung RTT for those service located in Africa. In terms of performance, the numbers are clearly in favor of this decision.

6.3 DNS performance

Given the importance of the server IP address location to reduce latency, DNS resolution plays an important role. For this, we drill down on DNS resolver choice and performance. We consider DNS/UDP traffic, for which we observe the original end-user device request and resolver response. First, we look for popular DNS resolvers, quantify their resolution latency, and next we observe the impact on server choice.

In total, we observe 4 195 of different resolvers, some of them only sporadically. Interestingly, we found that customers use well-know open resolvers instead of operator resolver, and strangely choose custom, unusual, and geographically distant resolvers. In Figure 10, we break down the top-8 resolvers in terms of volume, separated by country. For a given country, each column shows the percentage of DNS traffic for the different resolvers. In the rightmost column, we report the median response time observed at the ground station.

We note that the operator DNS (first row) is quite used only in European countries. In Ireland, Spain, and the U.K., it accounts for 44,%, 29,%, and 38 % of the DNS volume, respectively. With a mean resolution time of only 3.98 ms, it offers the best performance. As expected, Google DNS is popular everywhere. In Africa it resolves 86 % of requests in Congo and more than 50 % in the other African countries. Other popular open resolvers, namely CloudFlare DNS and OpenDNS, have a different popularity, usually below 10 %. The resolution time for all of them is on the order of 20 ms.

Nigeria is a peculiar case. We find that 12 % of traffic goes to a local Nigerian operator resolver. For this resolver, the ground RTT artificially inflates the resolution time to about 120 ms since packets have to travel from the ground station in Italy to Nigeria and back.

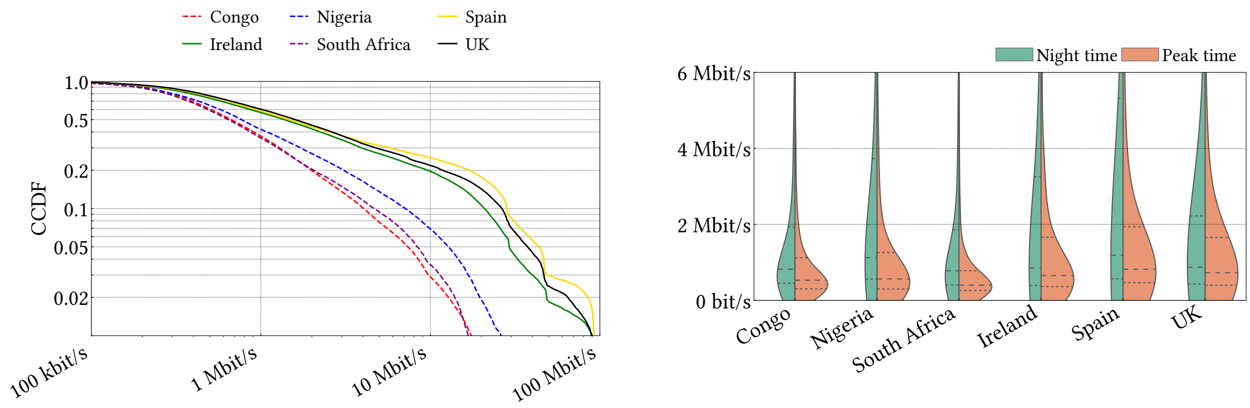
Interestingly, we observe two Chinese DNS resolvers (Baidu DNS and 114 DNS) in African countries, confirming the assumption that there is a significant Chinese community that use homeland resolvers. For Baidu, the resolution time is terrible, with a median response time higher than 350 ms that have to be added to the satellite RTT to reach the actual end-user device.

In summary, in most cases SatCom customers do not adopt the operator DNS and resort instead to open resolvers. Due to the particular routing in the SatCom network, we observe cases of resolvers that suffer very high RTT, yet they are widely used in Africa. This greatly impacts the DNS response time and introduces an additional 100 − 300 ms delay on top of the satellite RTT. This has a clear negative impact on the user experience.

6.4 Implications on server selection policies of CDNs and DNS resolvers

The superposition of i) routing constraints through the same ground station in Italy, ii) an intercontinental service presence that includes African and European countries, iii) the adoption of different DNS resolvers by customers, creates a very tangled picture that complicates the server selection policies of different CDN and DNS resolver operators. To examine these implications, we observe whether there are differences in ground RTT to the same service when using different DNS resolvers.

Table 2 shows some examples. We report the average ground RTT for some sample domains and some of the most popular resolvers for Nigeria and U.K. A more complete version of this table can be found in Appendix B. For U.K. (and European countries in general), the DNS resolver has little impact on performance. This is expected since i) the ground station is located in Europe and ii) customers tend to access European services that are well served by CDNs in Europe.



(a) Per country.

(b) Per time of the day.

Figure 11: Download speeds per customer.

However, this is not the case for African countries such as Nigeria. For example, the server IP address resolved to serve the `captive.apple.com` service results at 19.1 ms if resolved via the Operator-EU DNS for U.K. customers. It results instead at 110.4 ms if resolved by the 114DNS for customer in Africa. Interestingly, even the Google DNS resolver returns two different CDN nodes for U.K. (26.0 ms) and Nigerian (38.4 ms) customers. These resolvers provide more distant CDN IP addresses because they are likely confused by the originating customer request geo-position which conflicts with the routing through Italy. Not shown for brevity, we even observe that some DNS resolvers point to some CDN server in the original African country of the customer. This clearly inflates the ground RTT by several hundreds of milliseconds. Some domains, e.g., nflxvideo.net, are less affected by these phenomena. This may be because resolvers and CDNs have accurate information, or because they do not rely on DNS resolution to determine the closest CDN node, e.g., because they use Anycast-based CDN solutions (which are not affected by the DNS resolution issue).

A possible solution to the DNS inconsistency problem is to either force the use of the SatCom operator's resolver or work with the Open Resolver providers to correctly instruct the server selection policies to return the closest server to the ground station instead of the original location of the end user's terminals.

6.5 Throughput analysis

For the sake of completeness, we now briefly discuss download throughput. We measure it for TCP connections by calculating the gross ratio between bytes downloaded and the duration of the flow (calculated from the first to the last TCP segment with data sent).

Recall that we observe the TCP data flow from the Internet server to the ground station PEP. This download throughput is regulated by the actual download throughput from PEP to the end device, which happens to be the bottleneck. To obtain a reliable measurement, we only consider flows large enough for the throughput to reach stable values and to neglect the effects of buffering at the PEP.[9] For this purpose, we only consider flows carrying at least 10 MB of data, for which we limit the representations to 1 million samples from a three-day interval. Even in these cases, not all flows can be considered valid bulk download samples (e.g., persistent HTTP flows or rate-limited video streaming flows), and competing traffic may limit throughput. This figure can only be considered a rough estimate of the actual performance a customer gets.

Figure 11a shows the CCDF for download throughput separately by country. The operator offers several commercial plans with different maximum throughput. This is reflected in the knees of the curves in the figure. In Europe, where 30 Mb/s, 50 Mbit/s, and 100 Mbit/s plans are popular, we find that these customers can saturate their capacity with a single TCP flow. Overall, European countries have similar download throughput, with Ireland achieving slightly lower values due to its particular physical channel characteristics (see Section 6.1). For brevity, we limited our analysis to video streaming flows, and separated off-peak and on-peak times. In both cases, we could not find any signs of congestion.

In the African countries, the picture is quite different. First, the operator sells plans with a capacity of 10 Mb/s and 30 Mbit/s. Only few customers can saturate this capacity, with Nigerian customers tending to achieve slightly higher throughput. This is likely due to lower congestion on the satellite link. However, the higher congestion on the link, the less optimal server selection and routing, the presence of community WiFi APs, and likely the less powerful end-user terminals limit the maximum download throughput that customers can achieve.

This is confirmed by Figure 11b, which shows the distribution of download speed for each country, with nighttime hours separated from peak hours. Again, European customers have higher throughput than African customers. In all countries, throughput is lower during peak hours than at night, as shown in the body of the distribution and the lower percentiles and medians. The change is more pronounced in Congo and South Africa.

7 RELATED WORK

Since the Internet was born, characterization of network traffic has been an important topic to understand its evolution and usage.

[9]The PEP has a limited buffer per-user.

Seminal works analyze trends of Internet traffic trends in the 1990 s and early 2000 s [13, 14]. Since then, there has been a large body of research using passive measurements to understand Internet operations and usage patterns. Most works focus on characterizing Internet traffic generated by cable/fiber customers [3, 15, 25, 40], while others analyze traffic captured in the backbone network, mixing users with different access technologies [5, 23]. Recently, researchers have focused on analyzing the impact of the Covid-19 pandemic on Internet traffic, noting sudden and remarkable changes [6, 11, 12]. Our work is in this area and uses a similar methodology for collecting and analyzing passive measurements. We note that most measurements are related to Europe and North America, while we also provide insights into African Internet traffic, similar to Johnson *et al.* [19]. In addition, in this work, we have a unique opportunity to specifically study traffic from SatCom subscribers. To the best of our knowledge, we are the first to conduct a longitudinal study of SatCom traffic and present a comparison of traffic from different countries on two continents.

Satellite-based consumer Internet access [17] was launched in 2003 and has been the subject of a wide corpus of literature. Most of the work targets the design and optimization of communication channels, and a few studies successfully cover the topic [8, 21, 27]. Other work focuses on measuring and improving the performance of Internet protocols in SatCom. Tropea *et al.* [42] evaluates different TCP versions on geostationary satellite links, while Peng *et al.* [31] and Muhammad *et al.* [29] focus on the interaction of TCP and PEPs. More recently, the performance of QUIC and its interaction with PEPs have been studied [2, 22, 38]. For a complete benchmark of SatCom performance, see Deutschmann *et al.* [9], which provides figures on SatCom latency, throughput, and web page load time. Recently, the authors in [26] examined Starlink, the new low-orbit satellite communication system, which provides a comprehensive overview of the impact of this new communication on user-perceived performance when accessing globally distributed resources and the impact of different HTTP versions. Similarly, Mohamed *et al.* [24] propose a study from different vantage points to understand how the performance changes from the browser's perspective.

All of these works rely on active measurements, i.e., they deploy test environments with devices using SatCom and analyze the results. In this work, we provide a different perspective by providing performance data on SatCom through passive measurements by observing traffic from about 10 k customers and complementing the results of previous work.

8 CONCLUSION

In this work, we performed the first characterization of the SatCom network through passive measurements. By instrumenting the ground station of a major SatCom operator, where traffic from tens of thousands of customers is aggregated, we had the opportunity to characterize the traffic from different angles, providing a new perspective on a mature but complex technology.

Our results shed light on SatCom usage habits and traffic patterns, allowing us to compare Internet usage in different countries. We find that SatCom customers in Europe are more likely to be active during evening peak hours, while in Africa we observe a traffic peak in the morning. This suggests a classic leisure usage by European customers, while in Africa (at least partially) a business usage. Our measurements suggest that African subscriptions are used by multiple people. In fact, they consume much more chat services, which is hardly consistent with sole or domestic use. Customers can reach the nominal throughput of their plan and often access popular high-definition video streaming platforms. SatCom traffic reaches the Internet via a single ground station in Europe, which adds delays in accessing destinations in other continents. Nevertheless, we found that a non-negligible proportion of African customers use DNS resolvers in other continents.

ACKNOWLEDGMENTS

The research leading to these results has been funded by the SmartData@PoliTO Center for Big Data Technologies. We also acknowledge funding support from the Spanish State Research Agency under the project AgileMon (AEI PID2019-104451RB-C21) and from the Spanish Ministry of Science, Innovation and Universities under the program for the training of university lecturers (Grant number: FPU19/05678).

REFERENCES

[1] 2021. (How Much) Can Edge Computing Change Network Latency?. In *2021 IFIP Networking Conference (IFIP Networking)*. 1–9. https://doi.org/10.23919/IFIPNetworking52078.2021.9472847

[2] A Abdelsalam, Michele Luglio, Mattia Quadrini, Cesare Roseti, and Francesco Zampognaro. 2019. QUIC-proxy based architecture for satellite communication to enhance a 5G scenario. In *2019 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, 1–6.

[3] Zachary S Bischof, Fabián E Bustamante, and Rade Stanojevic. 2014. Need, want, can afford: Broadband markets and the behavior of users. In *Proceedings of the 2014 Conference on Internet Measurement Conference*. 73–86.

[4] J Border. 2001. Enhancing proxies intended to mitigate link-related degradations. *RFC 3135* (2001).

[5] Pierre Borgnat, Guillaume Dewaele, Kensuke Fukuda, Patrice Abry, and Kenjiro Cho. 2009. Seven years and one day: Sketching the evolution of internet traffic. In *IEEE INFOCOM 2009*. IEEE, 711–719.

[6] Timm Böttger, Ghida Ibrahim, and Ben Vallis. 2020. How the Internet reacted to Covid-19: A perspective from Facebook's Edge Network. In *Proceedings of the ACM Internet Measurement Conference*. 34–41.

[7] Conviva. 2021. Conviva's State of Streaming. https://pages.conviva.com/rs/138-XJA-134/images/RPT_Conviva_State_of_Streaming_Q3_2021.pdf.

[8] Joerg Deutschmann, Thomas Heyn, Christian Rohde, Kai-Steffen Hielscher, and Reinhard German. 2021. Broadband internet access via satellite: State-of-the-art and future directions. In *Broadband Coverage in Germany; 15th ITG-Symposium*.

[9] Jörg Deutschmann, Kai-Steffen Hielscher, and Reinhard German. 2019. Satellite internet performance measurements. In *2019 International Conference on Networked Systems (NetSys)*. IEEE, 1–4.

[10] Jinliang Fan, Jun Xu, and Mostafa H Ammar. 2004. Crypto-pan: Cryptography-based prefix-preserving anonymization. *Computer Networks* 46, 2 (2004).

[11] Thomas Favale, Francesca Soro, Martino Trevisan, Idilio Drago, and Marco Mellia. 2020. Campus traffic and e-Learning during COVID-19 pandemic. *Computer networks* 176 (2020), 107290.

[12] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, Oliver Hohlfeld, and Georgios Smaragdakis. 2020. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3419394.3423658

[13] Marina Fomenkov, Ken Keys, David Moore, and KC Claffy. 2004. Longitudinal study of Internet traffic in 1998-2003. In *Proceedings of the winter international synposium on Information and communication technologies*. 1–6.

[14] Chuck Fraleigh, Sue Moon, Bryan Lyles, Chase Cotton, Mujahid Khan, Deb Moll, Robert Rockell, Ted Seely, and S Christophe Diot. 2003. Packet-level traffic measurements from the Sprint IP backbone. *IEEE network* 17, 6 (2003), 6–16.

[15] José Luis García-Dorado, Alessandro Finamore, Marco Mellia, Michela Meo, and Maurizio Munafo. 2012. Characterization of isp traffic: Trends, user habits, and

access technology impact. *IEEE Transactions on Network and Service Management* 9, 2 (2012), 142–155.

[16] Jim Griner, John Border, Markku Kojo, Zach D. Shelby, and Gabriel Montenegro. 2001. Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations. RFC 3135. https://doi.org/10.17487/RFC3135

[17] Yurong Hu and V.O.K. Li. 2001. Satellite-based Internet: a tutorial. *IEEE Communications Magazine* 39, 3 (2001), 154–162. https://doi.org/10.1109/35.910603

[18] DPDK Intel. 2014. Data plane development kit.

[19] David L Johnson, Veljko Pejovic, Elizabeth M Belding, and Gertjan Van Stam. 2011. Traffic characterization and internet usage in rural Africa. In *Proceedings of the 20th international conference companion on World wide web*. 493–502.

[20] Thomas Koch, Ethan Katz-Bassett, John Heidemann, Matt Calder, Calvin Ardi, and Ke Li. 2021. Anycast In Context: A Tale of Two Systems. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference* (Virtual Event, USA) *(SIGCOMM '21)*. Association for Computing Machinery, New York, NY, USA, 398–417.

[21] Oltjon Kodheli, Eva Lagunas, Nicola Maturo, Shree Krishna Sharma, Bhavani Shankar, Jesus Fabian Mendoza Montoya, Juan Carlos Merlano Duncan, Danilo Spano, Symeon Chatzinotas, Steven Kisseleff, et al. 2020. Satellite communications in the new space era: A survey and future challenges. *IEEE Communications Surveys & Tutorials* 23, 1 (2020), 70–109.

[22] Nicolas Kuhn, François Michel, Ludovic Thomas, Emmanuel Dubois, and Emmanuel Lochin. 2020. QUIC: Opportunities and threats in SATCOM. In *2020 10th Advanced Satellite Multimedia Systems Conference and the 16th Signal Processing for Space Communications Workshop (ASMS/SPSC)*. IEEE, 1–7.

[23] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. 2010. Internet inter-domain traffic. *ACM SIGCOMM Computer Communication Review* 40, 4 (2010), 75–86.

[24] Mohamed M. Kassem, Aravindh Raman, Diego Perino, and Nishanth Sastry. 2022. A Browser-side View of Starlink Connectivity. In *Proceedings of the 2022 Internet Measurement Conference*. https://doi.org/10.1145/3517745.3561457

[25] Gregor Maier, Anja Feldmann, Vern Paxson, and Mark Allman. 2009. On dominant characteristics of residential broadband internet traffic. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*. 90–102.

[26] François Michel, Martino Trevisan, Danilo Giordano, and Olivier Bonaventure. 2022. A First Look at Starlink Performance. In *Proceedings of the 2022 Internet Measurement Conference*. https://doi.org/10.1145/3517745.3561416

[27] Daniel Minoli. 2015. *Innovations in satellite communications and satellite technology: the industry implications of DVB-S2X, high throughput satellites, Ultra HD, M2M, and IP*. John Wiley & Sons.

[28] Samuel Maredi Mojapelo. 2020. The internet access and use in public libraries in Limpopo Province, South Africa. *Public Library Quarterly* 39, 3 (2020), 265–282.

[29] Muhammad Muhammad, Matteo Berioli, and Tomaso De Cola. 2014. A simulation study of network-coding-enhanced PEP for TCP flows in GEO satellite networks. In *2014 IEEE International Conference on Communications (ICC)*. IEEE, 3588–3593.

[30] Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, and Zhi-Li Zhang. 2020. A first look at commercial 5G performance on smartphones. In *Proceedings of The Web Conference 2020*. 894–905.

[31] Fei Peng, Ángel Salamanca Cardona, Kaveh Shafiee, and Victor CM Leung. 2012. Tcp performance evaluation over geo and leo satellite links between performance enhancement proxies. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*.

[32] Alain Pirovano and Fabien Garcia. 2013. A new survey on improving TCP performances over geostationary satellite link. *Network and Communication Technologies* 2, 1 (2013), 1.

[33] Digital TV Research. 2022. Africa SVOD Forecasts. https://digitaltvresearch.com/product/africa-svod-forecasts/.

[34] Matthew Sargent and Mark Allman. 2014. Performance within a fiber-to-the-home network. *ACM SIGCOMM Computer Communication Review* 44, 3 (2014).

[35] Luca Schumann, Trinh Viet Doan, Tanya Shreedhar, Ricky Mok, and Vaibhav Bajpai. 2022. Impact of Evolving Protocols and COVID-19 on Internet Traffic Shares. *arXiv preprint arXiv:2201.00142* (2022).

[36] Laura Silver and Courtney Johnson. 2018. Internet connectivity seen as having positive impact on life in Sub-Saharan Africa. (2018).

[37] Christoph Stork, Enrico Calandro, and Alison Gillwald. 2013. Internet going mobile: internet access and use in 11 African countries. *info* (2013).

[38] Ludovic Thomas, Emmanuel Dubois, Nicolas Kuhn, and Emmanuel Lochin. 2019. Google QUIC performance over a public SATCOM access. *International Journal of Satellite Communications and Networking* 37, 6 (2019), 601–611.

[39] Martino Trevisan, Alessandro Finamore, Marco Mellia, Maurizio Munafo, and Dario Rossi. 2017. Traffic analysis with off-the-shelf hardware: Challenges and lessons learned. *IEEE Communications Magazine* 55, 3 (2017), 163–169.

[40] Martino Trevisan, Danilo Giordano, Idilio Drago, Maurizio Matteo Munafò, and Marco Mellia. 2020. Five years at the edge: Watching internet from the isp network. *IEEE/ACM Transactions on Networking* 28, 2 (2020), 561–574.

[41] Martino Trevisan, Ali Safari Khatouni, and Danilo Giordano. 2020. ERRANT: Realistic emulation of radio access networks. *Computer Networks* 176 (2020), 107289.

[42] M Tropea and P Fazio. 2013. Evaluation of TCP versions over GEO satellite links. In *2013 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*. IEEE, 86–90.

[43] Dongzhu Xu, Anfu Zhou, Xinyu Zhang, Guixian Wang, Xi Liu, Congkai An, Yiming Shi, Liang Liu, and Huadong Ma. 2020. Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption. In *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*. 479–494.

[44] J. Zhu, S. Roy, and J.H. Kim. 2006. Performance Modelling of TCP Enhancements in Terrestrial–Satellite Hybrid Networks. *IEEE/ACM Transactions on Networking* 14, 4 (2006), 753–766. https://doi.org/10.1109/TNET.2006.880167

A SERVICES

Table 3 reports the regular expressions defined to identify services and categories from domain names.

Table 4: Ground segment RTT for most popular second-level domains for Congo and South Africa per DNS resolver.

name	Congo									South Africa						
Resolver	114	Aliyun	Baidu	Cloudfare	Google	Level3	Norton	OpenDNS	Operator-EU	114	Aliyun	Baidu	Cloudfare	Google	OpenDNS	Operator-EU
apple.com	118, 3 ms	-	-	91, 3 ms	169, 8 ms	51, 0 ms	39, 9 ms	129, 4 ms	75, 4 ms	111, 3 ms	-	-	62, 6 ms	190, 6 ms	187, 3 ms	69, 0 ms
doubleclick.net	150, 6 ms	45, 9 ms	292, 1 ms	130, 2 ms	62, 7 ms	29, 3 ms	46, 4 ms	64, 4 ms	45, 4 ms	114, 5 ms	27, 5 ms	287, 4 ms	73, 9 ms	64, 9 ms	21, 1 ms	70, 3 ms
facebook.com	68, 1 ms	78, 9 ms	59, 8 ms	47, 2 ms	53, 4 ms	56, 5 ms	23, 1 ms	28, 6 ms	40, 2 ms	67, 6 ms	132, 9 ms	279, 5 ms	42, 6 ms	39, 7 ms	29, 3 ms	33, 6 ms
fbcdn.net	40, 6 ms	27, 8 ms	31, 9 ms	32, 1 ms	29, 7 ms	46, 3 ms	190, 8 ms	32, 2 ms	41, 9 ms	27, 4 ms	35, 1 ms	22, 6 ms	31, 4 ms	30, 1 ms	27, 7 ms	34, 3 ms
google.com	154, 0 ms	47, 5 ms	186, 0 ms	43, 1 ms	79, 8 ms	36, 5 ms	47, 5 ms	28, 1 ms	77, 9 ms	104, 6 ms	61, 2 ms	201, 3 ms	106, 3 ms	52, 9 ms	98, 2 ms	55, 3 ms
googleapis.com	165, 5 ms	42, 4 ms	323, 3 ms	66, 6 ms	77, 5 ms	54, 7 ms	32, 3 ms	31, 3 ms	74, 9 ms	111, 0 ms	32, 7 ms	275, 3 ms	66, 5 ms	64, 2 ms	28, 7 ms	79, 5 ms
googlevideo.com	145, 6 ms	34, 3 ms	69, 8 ms	95, 0 ms	87, 2 ms	37, 5 ms	63, 4 ms	25, 5 ms	81, 8 ms	24, 2 ms	30, 9 ms	86, 9 ms	35, 8 ms	38, 5 ms	84, 1 ms	48, 4 ms
gstatic.com	146, 8 ms	31, 2 ms	286, 1 ms	96, 6 ms	67, 9 ms	30, 8 ms	39, 1 ms	22, 9 ms	80, 2 ms	107, 6 ms	27, 5 ms	288, 5 ms	67, 8 ms	90, 3 ms	824, 3 ms	94, 7 ms
microsoft.com	118, 0 ms	-	281, 0 ms	59, 1 ms	101, 1 ms	96, 0 ms	77, 7 ms	97, 4 ms	77, 0 ms	106, 1 ms	-	-	46, 6 ms	107, 3 ms	107, 1 ms	72, 6 ms
netease.com	251, 2 ms	193, 2 ms	240, 3 ms	247, 9 ms	252, 0 ms	209, 4 ms	234, 8 ms	-	172, 1 ms	-	-	-	-	-	-	265, 4 ms
qq.com	266, 7 ms	254, 3 ms	269, 3 ms	269, 2 ms	251, 5 ms	255, 4 ms	233, 6 ms	254, 2 ms	243, 3 ms	-	-	253, 8 ms	115, 1 ms	253, 8 ms	-	243, 9 ms
scooper.news	120, 2 ms	50, 0 ms	-	97, 8 ms	132, 9 ms	29, 5 ms	36, 1 ms	23, 1 ms	36, 5 ms	-	-	-	1026, 0 ms	61, 0 ms	-	19, 9 ms
shalltry.com	82, 0 ms	40, 9 ms	43, 3 ms	95, 2 ms	68, 1 ms	83, 9 ms	66, 7 ms	37, 4 ms	69, 1 ms	-	-	-	42, 7 ms	41, 7 ms	-	39, 0 ms
tiktokcdn.com	215, 2 ms	-	207, 3 ms	124, 8 ms	93, 4 ms	29, 6 ms	100, 7 ms	69, 1 ms	53, 7 ms	137, 3 ms	-	-	56, 5 ms	173, 6 ms	129, 5 ms	73, 3 ms
tiktokv.com	202, 9 ms	24, 9 ms	280, 6 ms	94, 5 ms	190, 3 ms	50, 5 ms	59, 1 ms	114, 7 ms	54, 6 ms	89, 3 ms	-	-	41, 6 ms	193, 0 ms	186, 3 ms	48, 5 ms
umeng.com	307, 9 ms	106, 7 ms	302, 8 ms	300, 8 ms	242, 3 ms	252, 2 ms	262, 2 ms	82, 1 ms	239, 4 ms	346, 1 ms	-	-	267, 6 ms	140, 9 ms	-	104, 0 ms
whatsapp.net	70, 9 ms	54, 2 ms	85, 2 ms	59, 7 ms	51, 8 ms	47, 9 ms	47, 3 ms	23, 5 ms	44, 6 ms	67, 8 ms	134, 9 ms	76, 4 ms	38, 6 ms	42, 3 ms	20, 6 ms	29, 9 ms
yximgs.com	206, 6 ms	184, 7 ms	269, 4 ms	174, 6 ms	162, 7 ms	-	-	174, 2 ms	153, 9 ms	-	-	-	182, 8 ms	29, 9 ms	-	89, 2 ms

Table 3: Regular expressions used to identify services and service category.

Service	Regexp	Category
Spotify	[.spotify.com$, .scdn.co$, .scdn.com$]	Audio
Youtube	[.googlevideo.com$, .ytimg.com$, .youtube.com$, .gvt1.com$, .gvt2.com$, .youtube-nocookie.com$]	Video
Netflix	[.netflix., .nflxext., .nflximg., .nflxvideo., .nflxso.]	Video
Sky	[.sky.com$$]	Video
Primevideo	[amazonvideo.com$, primevideo.com$, pv-cdn.net$, atv-ps.amazon.com$, atv-ext.amazon.com$, atv-ext-eu.amazon.com$, atv-ext-fe.amazon.com$, atv-ps-eu.amazon, atv-ps-fe.amazon]	Video
Facebook	[.facebook.com$, .fbcdn.net$, .facebook.net$, ^fbcdn, ^fbstatic, ^fbexternal, .fbsbx.com$, fb.com$]	Social
Twitter	[.twitter., .twimg., ^twitter.com$, twitter.com.edgesuite.net, twitter-any.s3.amazonaws.com, twitter-blog.s3.amazonaws.com]	Social
Linkedin	[.linkedin.com$, .licdn.com$, .lnkd.in$]	Social
Instagram	[.instagram.com$, .cdninstagram.com$, ^igcdn]	Social
Tiktok	[tiktok.com$, tiktokcdn, tiktokv.com$]	Social
Google	[^www.google., ^google.]	Search engine
Bing	[.bing.com$]	Search engine
Yahoo	[.yahoo.com$, .yahoo.net$, .yimg.com$]	Search engine
Duckduck	[.duckduckgo.]	Search engine
Whatsapp	[.whatsapp.com$, .whatsapp.net$]	Chat
Telegram	[.telegram.org$, ^telegram.org$]	Chat
Snapchat	[.snapchat.com$, feelinsonice.appspot.com$, feelinsonice-hrd.appspot.com$, feelinsonice.l.google.com$]	Chat
Skype	[.skypeassets.com$, .skype.com$, .skype.net$]	Chat
Wechat	[wechat.com$, weixin.qq.com$, wxs.qq.com$]	Chat
Office365	[.sharepoint.com$, .office.net$, .onenote.com$, .office365.com$, .office.com$, teams.microsoft, teams.office, lync, skype, live.com$]	Work
Gsuite	[.googledrive.com$, .drive.google.com$, .docs.google.com$, .sheets.google.com$, .slides.google.com$, .talk.google.com$, .takeout.google.com$]	Work
Dropbox	[dropbox, db.tt$]	Work

B GROUND SEGMENT RTT

For the sake of completeness, Table 4 and Table 5 provide the average ground segment RTT for several countries and DNS resolvers.

Table 5: Ground segment RTT for most popular second-level domains for Nigeria and United Kingdom per DNS resolver.

name	Nigeria									United Kingdom						
Resolver	114	Aliyun	Baidu	Cloudfare	Google	Level3	OpenDNS	Operator-EU	Yandex	114	Aliyun	Baidu	Cloudfare	Google	OpenDNS	Operator-EU
apple.com	116, 9 ms	264, 9 ms	-	35, 7 ms	76, 6 ms	51, 8 ms	64, 1 ms	65, 0 ms	56, 0 ms	-	-	-	40, 6 ms	70, 7 ms	54, 1 ms	45, 1 ms
doubleclick.net	153, 2 ms	31, 8 ms	287, 8 ms	28, 9 ms	91, 6 ms	29, 3 ms	21, 6 ms	41, 2 ms	43, 3 ms	-	-	284, 4 ms	24, 0 ms	23, 3 ms	23, 9 ms	26, 9 ms
facebook.com	65, 2 ms	41, 7 ms	67, 4 ms	48, 1 ms	44, 9 ms	44, 0 ms	32, 3 ms	42, 2 ms	75, 8 ms	-	-	33, 9 ms	24, 8 ms	20, 5 ms	19, 8 ms	25, 7 ms
fbcdn.net	43, 7 ms	39, 2 ms	25, 4 ms	40, 2 ms	35, 4 ms	24, 2 ms	32, 7 ms	45, 0 ms	71, 9 ms	-	19, 1 ms	21, 7 ms	24, 4 ms	23, 3 ms	26, 3 ms	21, 0 ms
google.com	147, 9 ms	31, 2 ms	162, 2 ms	25, 8 ms	75, 6 ms	175, 5 ms	27, 6 ms	72, 3 ms	151, 9 ms	-	-	125, 6 ms	27, 0 ms	38, 0 ms	20, 7 ms	23, 1 ms
googleapis.com	164, 2 ms	34, 6 ms	287, 2 ms	25, 0 ms	71, 9 ms	29, 7 ms	22, 9 ms	95, 0 ms	136, 4 ms	-	45, 3 ms	273, 0 ms	23, 2 ms	154, 9 ms	22, 3 ms	25, 5 ms
googlevideo.com	106, 4 ms	34, 8 ms	64, 2 ms	24, 3 ms	55, 2 ms	45, 9 ms	28, 3 ms	78, 9 ms	113, 2 ms	-	19, 7 ms	293, 0 ms	22, 2 ms	24, 1 ms	18, 3 ms	23, 7 ms
gstatic.com	147, 5 ms	29, 8 ms	299, 8 ms	25, 9 ms	101, 2 ms	663, 2 ms	23, 6 ms	94, 0 ms	40, 2 ms	-	24, 0 ms	-	24, 0 ms	20, 5 ms	19, 2 ms	18, 7 ms
microsoft.com	169, 3 ms	-	-	-	93, 2 ms	90, 0 ms	79, 7 ms	77, 4 ms	46, 9 ms	-	-	-	56, 4 ms	56, 7 ms	57, 3 ms	36, 6 ms
netease.com	245, 4 ms	227, 6 ms	286, 0 ms	254, 3 ms	249, 2 ms	231, 4 ms	201, 5 ms	179, 7 ms	273, 8 ms	-	-	-	-	-	-	-
qq.com	245, 2 ms	327, 4 ms	282, 2 ms	268, 5 ms	244, 7 ms	94, 4 ms	226, 5 ms	238, 9 ms	90, 6 ms	-	-	-	-	-	-	-
scooper.news	101, 5 ms	-	-	25, 2 ms	25, 9 ms	30, 6 ms	21, 1 ms	40, 2 ms	140, 2 ms	-	-	-	-	-	-	-
shalltry.com	83, 3 ms	-	-	42, 0 ms	97, 7 ms	41, 3 ms	42, 9 ms	91, 4 ms	93, 7 ms	-	-	-	-	-	-	-
tiktokcdn.com	299, 5 ms	-	-	30, 7 ms	84, 3 ms	35, 0 ms	87, 7 ms	73, 2 ms	47, 4 ms	-	257, 9 ms	243, 5 ms	-	30, 5 ms	31, 1 ms	-
tiktokv.com	177, 5 ms	-	-	40, 4 ms	111, 4 ms	43, 4 ms	101, 3 ms	56, 2 ms	56, 6 ms	-	-	273, 2 ms	-	31, 7 ms	28, 2 ms	-
umeng.com	305, 1 ms	43, 2 ms	301, 5 ms	328, 8 ms	216, 3 ms	296, 2 ms	201, 6 ms	251, 8 ms	216, 1 ms	-	-	-	-	-	-	-
whatsapp.net	63, 7 ms	65, 8 ms	119, 4 ms	23, 6 ms	59, 7 ms	58, 7 ms	56, 4 ms	51, 3 ms	49, 0 ms	-	-	-	23, 3 ms	52, 6 ms	14, 0 ms	26, 2 ms
yximgs.com	226, 8 ms	36, 8 ms	263, 5 ms	-	157, 9 ms	-	-	180, 9 ms	-	-	-	-	-	-	-	-