An Open Dataset of Operational Mobile Networks

An Open Dataset of Operational Mobile Networks

Ali Safari Khatouni
Western University
London, Canada
asafarik@uwo.ca

Martino Trevisan
Politecnico di Torino
Torino, Italy
martino.trevisan@polito.it

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

Mohammad Rajiullah
Karlstad University
Karlstad, Sweden
moharaji@kau.se

Stefan Alfredsson
Karlstad University
Karlstad, Sweden
stefan.alfredsson@kau.se

Anna Brunstrom
Karlstad University
Karlstad, Sweden
anna.brunstrom@kau.se

Cise Midoglu
Simula Research Laboratory
Oslo, Norway
cise@simula.no

Özgü Alay
University of Oslo & SimulaMet
Oslo, Norway
ozgua@uio.no

ABSTRACT

Mobile networks have become ubiquitous and the primary means to access the Internet, and the traffic they generate has rapidly increased over the last years. The technology and service diversity in mobile networks call for extensive and accurate measurements to ensure the proper functioning of the networks and rapidly spot impairments. However, the measurement of mobile networks is complicated by their scale, and, thus, expensive, especially due to the diversity of deployments, technologies, and web services. In this paper, we present and provide access to the largest open international mobile network dataset collected using the MONROE platform spanning six countries, 27 mobile network operators, and 120 measurement nodes. We use them to run measurements targeting several web services from January 2018 to December 2019, collecting millions of TCP and UDP flows using these commercial mobile networks. We illustrate the data collection platforms and describe some of the main experiments. Besides a high-level overview of the dataset, we provide two practical use cases. First, we show how our data can be used as a proxy for web service performance. Second, we study the content delivery infrastructure of Facebook.

CCS CONCEPTS

• **Networks** → **Network experimentation**; **Network measurement**; **Mobile networks**; *Network monitoring*.

KEYWORDS

Open Dataset; Mobile networks; Web Services; Quality of Service

ACM Reference Format:
Ali Safari Khatouni, Martino Trevisan, Danilo Giordano, Mohammad Rajiullah, Stefan Alfredsson, Anna Brunstrom, Cise Midoglu, and Özgü Alay. 2020. An Open Dataset of Operational Mobile Networks. In *18th ACM International Symposium on Mobility Management and Wireless Access (MobiWac'20), November 16–20, 2020, Alicante, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3416012.3424619

1 INTRODUCTION

In the last decades, mobile networks have become ubiquitous, with pervasive mobile infrastructures allowing people to stay connected without the need for wired connections. The traffic generated by mobile networks is expected to have a growth of 46% in the period 2017-2022 [7], and smartphone-based services have recently gained unprecedented popularity [32]. This has been made possible by the advances in telecommunications with faster transfer speed and reduced latency. The 5G technology is expected to improve network performance once again and promises substantial cost savings, further fostering the spread of mobile networks. In this context, the quick boom (and vanish) of web services and deployment of new technologies (e.g., encryption [18]) makes it essential to run extensive and accurate measurements with the goal of ensuring the proper functioning of the networks and rapidly spotting anomalies and impairments. However, the measurement of mobile networks is complicated by their scale, the diversity of deployments and technologies, and, finally, a large number of services that users access.

In this paper, our goal is to share a large dataset obtained on a one-of-a-kind monitoring infrastructure spanning 6 countries with a considerable amount of experiments and 27 Mobile Network Operators (MNOs) [1]. To this end, we present the data collected in the MONROE project, in which 120 nodes have been instrumented to run measurements accessing millions of web pages over a period of two years from January 2018 to December 2019. The scale of the MONROE project and the duration of the experiments allow us to catch the diversity intrinsic to mobile networks and study

[1]The dataset is available here: https://smartdata.polito.it/mobile-networks-open-dataset

the infrastructure of most popular web services and operators. In addition to the effort of engineering the infrastructure and making the experiments, we focus on the delicate operation of storing, merging, and aggregating data coming from heterogeneous sources and distant locations. Indeed, the data we publish report flow and application-level measurements that the MONROE users actively scheduled. We augment them with rich contextual information passively recorded regarding the physical conditions and geographical location of the mobile nodes.

Besides providing a thorough characterization of our dataset, we show two practical use cases where our measurements can be used as a proxy for network performance or improve the understanding of web-service infrastructure. We strongly believe that open datasets are of crucial importance for researchers to fully understand the Internet. Moreover, large datasets, collected in a realistic environment are a prerequisite for training reliable machine learning models and artificial intelligence applications in general [6]. As such, we reckon our data will be helpful for the research community and network practitioners in general.

The paper is organized as follows. Section 2 presents related works. Section 3 illustrates the measurement platform and experimental campaigns. Section 4 introduces the dataset description. Section 5 presents the two use cases. Finally, Section 6 concludes the paper.

2 RELATED WORK

This paper contributes by offering to the community the dataset collected during the MONROE project. However, we are not the first to release an open dataset on network traffic. Given the drop in memory cost, the rise of internet traffic network monitoring is becoming a fundamental task for network operators, researchers, and network managers to understand the behavior of the network and users' QoE. In this direction, nationwide [11, 29, 32] and international [4, 24] efforts have been put to monitor network performance. In this direction, several works address specific network topics, Ren *et al.* [26] and Sivanathan *et al.* [28] propose datasets to study the behavior of Internet of Things devices in testbed environments. The MAWI (Measurement and Analysis of the Wide Internet) group [26] offers network traces including different applications and network conditions, describing the presence of various known anomalies. Different from the previous works, the MONROE datasets are collected as a wide range of measurements from real operation networks in different countries. The Monroe datasets have been already extremely useful to evaluate the performance and evolution of Mobile Networks. Safari *et al.* [17, 27] illustrate the utility of such a dataset on evaluating the variation of mobile network performance. They show how network performance is highly unpredictable due to the dynamicity of the mobile networks. Rajiullah *et al.* [25] exploit more than two million page visits to show the impact on web performance of different web protocols, browsers, and access technologies. By having access to different mobile nodes of the same operator in different countries, Mandalari *et al.* [21] show the impact of roaming connectivity on the network traffic and experienced performance. Recently, Trevisan *et al.* [33] create a data-driven mobile network emulator based on MONROE datasets showing the importance of the measurement campaigns to create accurate tools.

This dataset differs from previous datasets published by MONROE consortium. The previous papers just published a specific measurement campaign and aggregated dataset, i.e., a limited number of Tstat metrics. This dataset is the only one with the whole two years of data and the full Tstat metrics and metadata. To the best of our knowledge, we are the first to make public such large data from real operational networks in different countries. Given the complexity of collecting and sharing data, we believe that despite being used in the aforesaid works, the MONROE datasets could be useful for the research community in many different contexts.

3 MEASUREMENTS PLATFORM AND EXPERIMENT CAMPAIGNS

3.1 MONROE platform

The MONROE platform [4, 20] has been used to collect the dataset. MONROE is a unique open-source platform to run measurement campaigns in 4 countries in Europe (Italy, Norway, Spain, Sweden). It consists of more than 120 nodes deployed in stationary or mobile scenarios. The measurement node is the core of the platform. It consists of two main components, namely the hardware configuration, and the software ecosystem. Each node is equipped with at least three LTE modems, with commercial data plan subscriptions with mobile operators. Figure 1 shows an overview of the MONROE platform setup. Some of the stationary nodes are also equipped with an Ethernet connection. The node software is based on a Linux Debian "stretch" distribution to ensure long term stability. The nodes run the management and maintenance software with the experimentation enablers. The experiments run as Docker containers to enable a consistent experiment environment across nodes and developer platforms. Running experiments in docker containers also provide isolation between experiments themselves, and the node operating system. The results of the experiments are transferred periodically from the nodes to a remote repository for further processing, archiving, and insertion into a Cassandra database to enable data queries and analysis. There are two groups of open call users [2] who have used the platform to run a wide range of experiments [16]. In addition to them, several large-scale measurement campaigns have been run by the consortium. Therefore, we have measurements from 4 countries during the whole 2 years, roaming measurement involves 2 additional countries (the United Kingdom and Germany, i.e., the total comes to 6 countries), and some more external users countries. In a nutshell, the dataset has a considerable number of experiments in these 6 countries (Italy, Norway, Spain, Sweden, the United Kingdom, and Germany). We further detail the main experiment campaigns in the following.

3.1.1 Metadata. MONROE nodes augment the information about the access link status. The MONROE metadata [3] are event-based data collected by passively monitoring the statistics exposed directly from the modems through their management interface. It is the full context information about the state of a node, e.g., visited network

[2] https://www.monroe-project.eu/opencall1/
https://www.monroe-project.eu/opencall2/
[3] https://github.com/MONROE-PROJECT/Experiments/wiki/Metadata

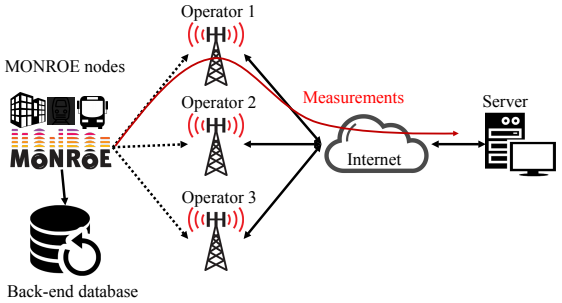


Figure 1: Overview of the MONROE platform and measurements.

Mobile Country Code (MCC), Mobile Network Code (MNC), signal strength, frequency, technology in use, cell-ID, etc., and its location as from GPS. The metadata broadcasting service runs continuously in the background and relays metadata to experiment containers.

3.1.2 Tstat. Each MONROE node runs Tstat [31], a passive meter that exports rich log files regarding all the traffic flowing from the nodes to the Internet. Tstat generates only flow records, i.e., a single entry for each TCP/UDP stream with per-flow metrics. Streams are evicted either by the observation of particular packets (e.g., TCP packets with RST flag set) or by a timeout. Each record contains classical fields on flow monitoring [15], such as server and client IP addresses and port numbers, packet, and byte-wise counters. Additional modules extract fields from packet payloads, such as the information seen in the Application-Layer Protocol Negotiation (ALPN) fields of TLS handshakes, which allows us to identify HTTP/2 and SPDY flows, and fields from QUIC public headers.

Tstat also indicates the *domain name* of contacted servers, which allows us to identify the service the node is contacting [30]. Tstat first searches the information in HTTP `Host:` headers and in the TLS Server Name Indication (SNI) within TLS Client Hello messages. For HTTP/TLS flows missing such fields and for other protocols (e.g., QUIC), Tstat exports the hostname the client resolved via DNS prior to open the flow. This is achieved by caching DNS traffic directed to any DNS resolver. Whenever a new flow is observed, Tstat searches for the last query performed by the same client that resulted in the contacted server IP address. This mechanism, called DN-Hunter, is explained in detail in [9], where it is shown that the association is correct for more than 90% of the flows.

The dataset we provide to the community includes all the Tstat log files generated on the MONROE nodes, which summarize all the traffic exchanged with the Internet. We enrich it with the contextual information describing the node status, especially regarding the operation of the network interfaces (signal strength, radio access technology, etc.).

3.1.3 Web browsing. For our web experiments, we leverage a customizable Docker container called MONROE-browsertime [1] to mimic a mobile device browser and retrieve mobile versions of visited pages. The work is based on [2], which we specifically engineered and deployed on MONROE nodes. We configured MONROE-browsertime to mimic a mobile device browser by setting both the screen resolution and the user-agent accordingly. The X virtual frame-buffer (Xvfb) [3] was used to provide the browser with a virtual display to render the web page since the MONROE nodes are running headless without a physical display. MONROE-browsertime provides a configurable experiment template to enable web measurements. We configure each measurement by controlling (i) the network to test (Ethernet, or the desired MBB interfaces), (ii) the browser (Firefox or Chrome), and (iii) the web protocol (HTTP/1.1 (H1) [12], HTTP/2.0 (H2) [8] or QUIC [19]). A combination of these parameters builds an *experiment setup.*

We select a list of target pages to visit. Given a network to test, we shuffle the order of pages to visit. Next, we visit each page with every browser and protocol combination, in random order. The visit of all pages with one network setup constitutes a *run.* Browser caching and cookies are active during each run and is reset after the last page has been tested. This ensures that each run starts from the same state, with a cold cache and reset cookies. Note that we use separate profiles for the same browser, one for each protocol, so that caches are separated (i.e., visits with H1 do not interfere with next visits to the same pages with H2 and vice-versa). MONROE-browsertime tracks a number of performance metrics such as Page Load Time (PLT), FirstPaint (FP), and RUMSpeedIndex (SI).

In addition to numerous test measurements, we ran two major web measurement campaigns on the MONROE platform. The first campaign ran from the 1st of April 2018 to the 4th of June 2018. This dataset focuses on measurements considering H1 and H2 protocols using both Firefox (version 56.0.1) and Chrome (version 64.0.3282.186). In this campaign, we selected 20 target pages on popular websites with H2 support from the most viewed sites in the Alexa [5] top ranking. We run a separate measurement campaign for measuring the performance of the QUIC protocol as compared to H2. This batch of experiments ran during a week, from the 5th to the 10th of June 2018. We focused here on websites that support QUIC (version HTTP/2+QUIC/39). We used only the Chrome browser since it was the only browser supporting QUIC.

Our second largest measurement campaign is running since July 2019. For this campaign, we selected 10 target pages, all of them with QUIC (version: HTTP/2+QUIC/43) support except for Facebook. We focus on pages that support QUIC to be able to use all three web protocols and include Facebook to capture its possible transition to QUIC (not happened yet). We again only use the Chrome browser (version: 71.0.3578.98) since Firefox still does not support QUIC. Readers are referred to [25] for more details on MONROE-browsertime.

3.1.4 MONROE-Nettest. MONROE-Nettest [23] is a configurable open-source tool built as an Experiment as a Service (EaaS) over the MONROE platform, which allows for conducting repeatable speedtest measurements in the complex mobile ecosystem. Its implementation is based on RMBT [4], a tool that is used by a number of National Regulatory Authorities (NRAs) in Europe for their crowdsourced network measurement applications. It is therefore compatible with many existing regulatory speedtest tools.

[4] https://www.netztest.at/doc/

MONROE-Nettest works in a client-server architecture, where the client initiates an active measurement session. The tool provides a lightweight client implementation using Docker virtualization. (1) the client container makes a test request to the measurement server, (2) the measurement server replies with the notion of availability, (3) the measurement, composed of 6 phases, is run between the client and the server, (4) the results are gathered at the client-side. This flow is easily applicable to testbeds, as well as highly scalable.

The measurement phases are Initialization, Pre-Test Downlink, Ping Test, Downlink Test, Pre-Test Uplink, and Uplink Test. Initialization consists of the client connecting to the measurement server and establishing the desired number of TCP flows. The pre-test phases are undertaken to ensure that the Internet connection is in an "active" state, i.e., that dedicated radio resources are available. The ping test consists of the client sending the desired number of TCP "ping"s in short intervals to the server to test the Round-Trip Time (RTT) of the connection. The client sends *n* small TCP packets (ASCII PING), to which the server replies with an ASCII PONG. The DL and UL tests are the main components of the measurement where within multiple TCP flows, the receiver side simultaneously requests and the sender side continuously sends data streams consisting of fixed-size chunks. After the nominal duration, the sender stops sending further chunks on all connections, the last chunk per each thread is allowed to transmit completely, and the DL/UL data rate of the connection is estimated. The MONROE-Nettest client is highly customizable with over 20 configuration parameters, further details can be found under [22].

We have scheduled MONROE-Nettest as a base experiment on all available nodes starting from July 2017, running 2, 3, and later 6 times a day. We have employed MONROE nodes as clients, and 2 well-provisioned MONROE-Nettest servers in Norway and Sweden, deployed on virtual machines hosted by the MONROE Alliance. The parameters we explicitly set for the base experiments were as follows: number of TCP flows 5, pre-test DL and pre-test UL duration 1s, DL and UL test duration 10s, number of pings 11. The TCP flows from these base experiments, as well as additional measurement campaigns using the MONROE-Nettest tools make up a part of the dataset provided in this work.

3.1.5 Roaming. The roaming measurement campaign [21] ran an extensive measurement to understand the roaming ecosystem in Europe after the "Roam like Home" initiative [5]. It opted to carry a series of measurements to study how mobile network operators implement roaming across Europe. There are 16 MNOs that support roaming in Europe involved in this measurement campaign. It consists of a collection of the above-mentioned measurements and some additional measurements to pinpoint the implications of roaming. It aims at identifying the roaming setup, infers the network configuration, and understands the impact on the user experience. It includes ping, traceroutes, DNS lookups, MONROE-browsertime, and VoIP traffic differentiation.

4 DATASET DESCRIPTION

In this section, we offer a quantitative overview of the dataset. Table 1 provides a breakdown of the collected data, showing different measures separately by the country. We also keep separate experiments done using mobile vs wired connection. Overall, in each country, we used 3-4 different MNOs. However, due to our roaming measurement campaign, we used these MNOs under different identities – i.e., we used SIM cards from 8-15 foreign operators on each country. Each node performed different sets of experiments, which caused them to contact 31-39 different web services and tens of thousands of servers. In total, they exchanged thousands of GB of data with the Internet consisting of millions of TCP flows. Considering countries, the traffic is concentrated in Sweden, Norway, Italy, and Spain. We observe less traffic for the other countries where MONROE deployed its nodes, and sporadic experiments are present for Poland, Portugal, and France.

The activity of nodes was not constant over time, as specific experimental campaigns required a high effort during a relatively short amount of time. In Figure 2, we show the trend of the generated traffic over time. The TCP traffic (Figure 2a) reaches its peak during the summer months of spring 2018, when, globally, the nodes generated 20 M TCP flows. Then it decreases and becomes finally stable during late 2019. More stable is the picture for UDP (Figure 2b). On average, each month the nodes generated tens of millions of UDP flows. Notice however that UDP mostly carries control traffic (DNS primarily) which accounts for a small traffic volume if compared with TCP. Finally, the figures provide separate lines for traffic exchanged while roaming on a foreign operator or using a local SIM card. Roaming is always present, to a lesser extent. In particular, during the first months of 2019, it almost vanishes (especially for TCP), meaning that few experiments were set up. However, during the peak months in spring 2018 [21], each month the nodes issued almost 10 M flows, summing TCP and UDP.

Finally, we characterize how each service is pervasive in our dataset. For this, we rely on the regular expression developed in [32] to map the contacted domain name into the name of the contacted service. Based on this we get the information about 41 different services plus 1 service (other) containing all the domain name for which we do not have a map. In Figure 3 we report how many flows involve each service in every month of our dataset. Services are sorted according to the number of months in which we observe flows toward that service. Top 10 services are the more pervasive ones having flows almost in every month for which we have data. Looking the type of the services we can see that we have a good repreentation of types such as: social network (i.e., Facebook, twitter, linkedin, and Instagram), search engine (i.e., Google, Yahoo, and Bing),entertainment (YouTube, bbc, and adult contents). This highlights how other the spacial heterogeneity given by the spacial and network operator diversity, the MONROE dataset offer a wide visibility of different type of services on which researchers may be interested in.

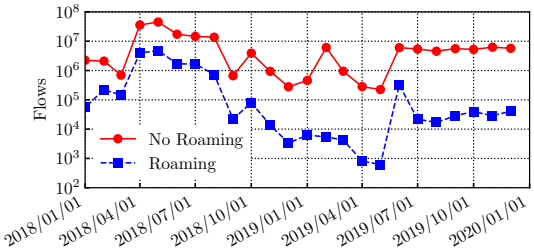
5 USE CASES

In this section, we briefly illustrate two possible use cases that demonstrate the utility of our data. We first provide some figures on the service performance that we can extrapolate from the TCP-level measurements. Then, we show how we can study the content-delivery infrastructures, focusing on the Facebook case.

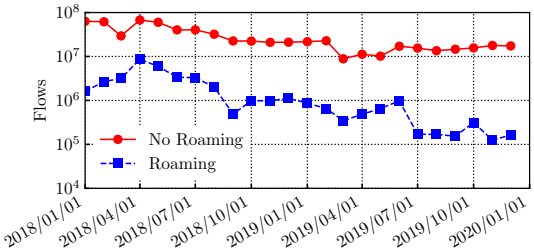
[5] https://ec.europa.eu/digital-single-market/en/roaming

Table 1: An overview of the dataset.

Country	# Nodes	# Distinct HNI	# Services	# Distinct contacted IP	# months	Total Download (GB)	Total Upload (GB)	# TCP flows (M)
Sweden	56	15	39	55939	24	11493	2902	78
Sweden	29	Wired	33	15090	23	239	32	2
Norway	33	13	39	44041	24	6286	1453	56
Norway	24	Wired	31	14706	23	224	20	3
Italy	25	15	37	49102	24	2234	120	33
Italy	16	Wired	33	13351	18	185	10	1.6
Spain	11	17	37	38393	24	1413	74	20
Spain	18	Wired	32	6985	16	286	22	0.9
United Kingdom	2	8	31	10321	8	17	0.83	0.3
United Kingdom	3	Wired	3	105	5	0.06	< 0.01	< 0.01
Germany	1	11	31	14502	7	36	1.68	0.7
Germany	2	Wired	1	3	3	< 0.01	< 0.01	< 0.01



(a) TCP.



(b) UDP.

Figure 2: The total number of flows over time for Roaming and Non-Roaming. Notice the log scale on the y-axis.

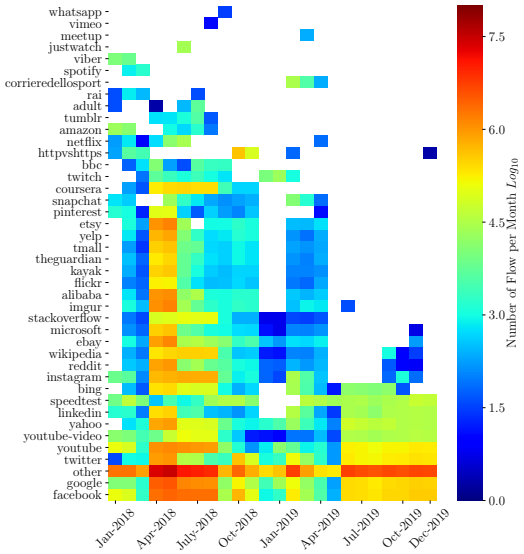


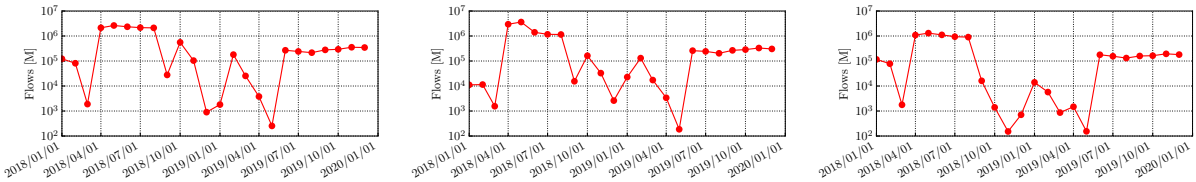
Figure 3: The number of flows for services over time. Services are sorted based on the number of months in the dataset.

5.1 Service Performance

The first use case we show focuses on the performance of the web services. Indeed, the measurement campaigns that build our dataset include first of all visits to websites. Here, we want to provide some figures that describe how their performance varies depending on the user's location and over time. Measuring it using passive measurements is not trivial, but some network-level metrics have been shown to be correlated with users' QoE [10]. In general, Tstat logs are annotated with a few QoS metrics that we can use to have a qualitative estimation of the performance of the services under study. Here, we focus on two QoS-related metrics: Round-Trip Time (RTT) and Time To First Byte (TTFB), which can be found in Tstat logs, and are, also intuitively, linked to the performance of web browsing.

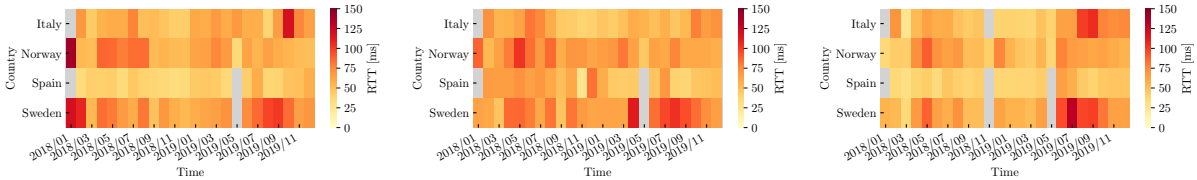
In this section, we focus on three web services for which we have sizeable and continuous measurements: We consider Facebook, Google, and YouTube. Figure 4 shows the number of flows captured in our dataset for each month of 2018 and 2019. Comparing it with Figure 2, we note that these measurements follow the general trend of the MONROE nodes activity. In particular, during spring 2019 we observe fewer measurements, but we seldom find less than 1000 k flows in a month.

In Figure 5, we represent in the form of heatmap the RTT observed by Tstat, separately by country and month. Three sub-figures show results for the three services under study. Tstat computes the RTT measuring the time between TCP data segments from the clients and the respective acknowledgment from the server, and each cell on the heatmap reports the *median* value across multiple measurements. A gray cell indicates that our dataset includes less than 100 flows, and, as such, we could not compare a reliable statistic. Looking at the figure, it is possible to have a general overview



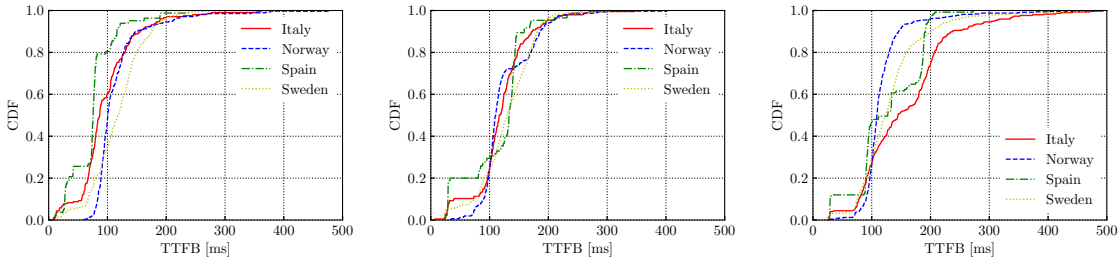
(a) *Facebook.* (b) *Google.* (c) *YouTube.*

Figure 4: The number of flows over time.



(a) *Facebook.* (b) *Google.* (c) *YouTube.*

Figure 5: Average RTT.



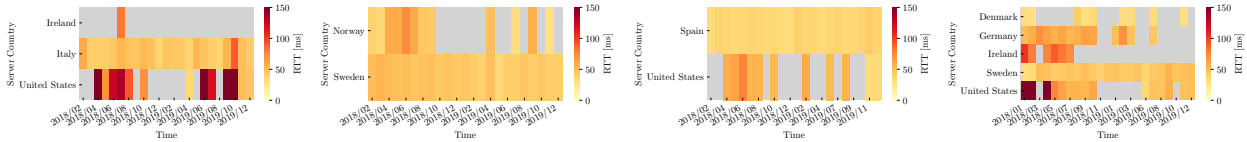
(a) *Facebook.* (b) *Google.* (c) *YouTube.*

Figure 6: Distribution of TTFB.

(a) Italy. (b) Norway. (c) Spain. (d) Sweden.

Figure 7: Word cloud presentation of the percentage of the total amount of data from Facebook servers' countries for different countries separately.

of a vast geographical and temporal span. In general, the figure confirms that large tech companies deploy edge servers close to users. Indeed, we observe a large number of cells which indicate a median RTT of 20 ms or below. However, the picture is not flat, and we observe significant variations over time. During some months, the RTT increases significantly, revealing either network impairments or changes in content/server location by the service. It is particularly interesting the case of YouTube, which shows a sizeable increase of RTT during summer/fall 2019 when it increases up to $\approx$ 100 ms contemporary in both Italy and Sweden. Studying the root causes of these phenomena is hard [13] and out of the scope of this paper. However, we argue that our dataset is useful to discover the situation of performance degradation which may hamper the



(a) Italy. (b) Norway. (c) Spain. (d) Sweden.

Figure 8: Temporal evolution of RTT for different Facebook servers country.

Table 2: The overall content geolocation statistics for Facebook. RTT, TWHT, TTFB, D are in ms, ms, ms, and GB, respectively.

	Italy					Norway					Spain					Sweden				
Country	**RTT**	**TWHT**	**TTFB**	**D**	%	**RTT**	**TWHT**	**TTFB**	**D**	%	**RTT**	**TWHT**	**TTFB**	**D**	%	**RTT**	**TWHT**	**TTFB**	**D**	%
Italy	38	48	87	1141	95.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Norway	-	-	-	-	-	49	38	103	327	22.5	-	-	-	-	-	-	-	-	-	-
Spain	-	-	-	-	-	-	-	-	-	-	38	37	72	789	97.4	-	-	-	-	-
Sweden	-	-	-	-	-	47	46	89	1119	76.7	-	-	-	-	-	40	42	118	2873	93.6
Denmark	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32	36	80	53	1.8
United States	48	143	287	48	4.1	-	-	-	-	-	58	53	131	14	1.7	153	153	326	54	1.8
Netherlands	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	54	51	112	46	1.5
Germany	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46	47	118	36	1.2

users' QoE due to the sudden increase of the RTT [14] and can be used to confirm (or contradict) well-known trends.

We complement the above analysis with Figure 6, which shows the distribution of the TTFB for the three services as observed by Tstat. Here, we want to compare different countries, limiting the variability introduced by the temporal dimension. As such, we use data only for the months of April and May 2018, when we collected millions of flows per each service. The TTFB measures the time elapsed since the beginning of a TCP connection to the first data segment from the server. Intuitively, it is the time taken to start receiving the response object in an HTTP transaction.[6] Looking at the figures, we first notice that the TTFB distribution is often multi-modal. For instance, for Google in Spain (green dashed line in Figure 6b) we observe TTFB of $\approx$ 30 ms in 20% of the cases, while around 140 ms in the majority of cases. Further manual checks reveal that this behavior is the effect of the placements of the CDN nodes of Google, which are very close in terms of RTT to some nodes, while for others the distance is significantly higher, even within the same country. In general, we observe that the TTFB is typically in the order of $100 - 200$ ms, with some privileged cases in which the performance is significantly better.

5.2 Content Location

In this section, we present the content geolocation distributions in the dataset. Here, we aim to provide an example of the usefulness and richness of the collected dataset. This work does not focus on the accurate geo-locating of an IP address. We rely on open-access databases[7] to locate an IP address. We present proximity of the content for the service with the highest volume of traffic, i.e., Facebook. A similar conclusion can be derived for other services but it is omitted for the sake of brevity.

Figure 7 presents the word cloud of the contacted servers country accessed by nodes when accessing Facebook service, illustrating the percentage of the total downloaded data that is retrieved from that country. It is fairly distinguishable that more than 90% of content is delivered by a server in the same country as the node (local server). Figure 7b reports the only exception in Norway, where 76% of the data volume goes across the border and comes from Sweden.

Figure 8 shows the temporal evolution of the RTT for the country of origin (y-axis) for the four countries separately. The gray cells indicate times without samples for that country of origin. The RTT values in the figure are calculated as median values over all samples for all operators in the country for a Facebook visit. The temporal evolution demonstrates two important observations. Firstly, the content provided by the local server experience lower RTTs (brighter cells). Typically, these giant content providers in today's Internet achieve this by having caches in most countries or using Content Delivery Networks (CDN). Secondly, the content from the same country is continuously provided by these local servers. Figure 8b reflects the continuous content delivery from Sweden to Norwegian nodes.

Finally, Table 2 presents a summary of the characteristics observed in the 4 countries. The first column of the table for each country indicates the median RTT for each origin servers' country. Column 2 reports the median value observed for TCP connection establishment time, Three-Way Handshake Time (TWHT). The third column shows the median TTFB. Columns 4 and 5 show the total amount of data provided by the servers in a country and its percentage of the total, respectively. Table 2 demonstrates that the content is primarily provided by the local servers, which have better connection characteristics.

We present content geolocation for a specific service in four countries. This is an example to show the richness of the dataset. The main limitation of these results are as follow: These results are based on open source IP location databases. We also use the most recent version of the database (2020), while the data contain measurements for 2018 and 2019 which can bias our analysis. However, we do not further discuss and validate the accuracy of these databases which is out of the scope of this paper.

[6] We are neglecting the TLS handshake in case of HTTPS flows.
[7] MAXMIND: https://dev.maxmind.com/geoip/geoip2/geolite2/ and IP2Location: https://lite.ip2location.com/database/ip-country

6 CONCLUSION

Mobile networks have turned into the most popular and easy way to access the Internet. They nowadays account for a large portion of the total Internet traffic with a continuous increase. Their diverse nature and technological heterogeneity call for large-scale measurement to ensure the proper functioning of the networks and anomaly detection. In this paper, we presented and provided access to the largest open international mobile network dataset collected using the MONROE platform in six countries, 27 MNOs, and 120 nodes from January 2018 to December 2019. We illustrated the data collection platforms and described the major experiments. We then presented a high-level overview of the dataset with a thorough characterization of our dataset. Finally, we demonstrated the importance of the dataset with two practical use cases, i.e., the performance of web services and a study on the content-delivery infrastructures of Facebook.

ACKNOWLEDGEMENTS

This work was partly funded by the *MONROE* Alliance, the SmartData@PoliTO center, and Region Värmland via the DigitalWell Arena project (Dnr RV2018-678).

REFERENCES

[1] [n.d.]. MONROE-Browsertime. https://github.com/MONROE-PROJECT/Experiments/. [Online; accessed 18-Jul-2020].
[2] [n.d.]. sitespeedio/browsertime: Your browser, your page, your scripts! https://github.com/sitespeedio/browsertime. [Online; accessed 20-Feb-2019].
[3] [n.d.]. Xvfb manual. https://goo.gl/EPBtpt. [Online; accessed 20-Feb-2019].
[4] Ö. Alay, A. Lutu, M. Peón-Quirós, V. Mancuso, T. Hirsch, K. Evensen, A. Hansen, S. Alfredsson, J. Karlsson, A. Brunstrom, A. Safari Khatouni, M. Mellia, and M. Ajmone Marsan. 2017. Experience: An Open Platform for Experimentation with Commercial Mobile Broadband Networks. *Proc. ACM MobiCom '17* (2017), 70–78. https://doi.org/10.1145/3117811.3117812
[5] Alexa. 2017. The top sites on the web. https://www.alexa.com/topsites. [Online; accessed 24-Jul-2020].
[6] Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Paolo Garza, Danilo Giordano, Marco Mellia, and Luca Venturini. 2016. SeLINA: A self-learning insightful network analyzer. *IEEE Transactions on Network and Service Management* 13, 3 (2016), 696–710.
[7] T Barnett, S Jain, U Andra, and T Khurana. 2018. Cisco Visual Networking Index (VNI), Complete Forecast Update, 2017–2022. *Americas/EMEAR Cisco Knowledge Network (CKN) Presentation* (2018).
[8] M. Belshe, R. Peon, and M. Thomson. 2015. Hypertext Transfer Protocol Version 2 (HTTP/2). RFC 7540 (Proposed Standard).
[9] Ignacio Bermudez, Marco Mellia, Maurizio M. Munafò, Ram Keralapura, and Antonio Nucci. 2012. DNS to the Rescue: Discerning Content and Services in a Tangled Web. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC'12)*. 413–426. http://doi.acm.org/10.1145/2398776.2398819
[10] Enrico Bocchi, Luca De Cicco, and Dario Rossi. 2016. Measuring the Quality of Experience of Web users. *ACM SIGCOMM Computer Communication Review* 46, 4 (2016), 8–13.
[11] FCC. 2013. *2013 Measuring Broadband America February Report.* Technical Report. FCC's Office of Engineering and Technology and Consumer and Governmental Affairs Bureau.
[12] R. Fielding et al. 1999. Hypertext Transfer Protocol – HTTP/1.1. RFC 2616 (Draft Standard).
[13] Danilo Giordano, Stefano Traverso, Luigi Grimaudo, Marco Mellia, Elena Baralis, Alok Tongaonkar, and Sabyasachi Saha. 2015. YouLighter: A cognitive approach to unveil YouTube CDN and changes. *IEEE Transactions on Cognitive Communications and Networking* 1, 2 (2015), 161–174.
[14] Danilo Giordano, Stefano Traverso, Luigi Grimaudo, Marco Mellia, Elena Baralis, Alok Tongaonkar, and Sabyasachi Saha. 2015. Youlighter: An unsupervised methodology to unveil youtube cdn changes. In *2015 27th International Teletraffic Congress*. IEEE, 19–27.
[15] Rick Hofstede, Pavel Čeleda, Brian Trammell, Idilio Drago, Ramin Sadre, Anna Sperotto, and Aiko Pras. 2014. Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX. *Commun. Surveys Tuts.* 16, 4 (2014), 2037–2064.
[16] Ali Safari Khatouni, Marco Ajmone Marsan, Marco Mellia, and Reza Rejaie. 2018. Deadline-constrained content upload from multihomed devices: Formulations and algorithms. *Computer Networks* 142 (2018), 76 – 92. https://doi.org/10.1016/j.comnet.2018.06.008
[17] Ali Safari Khatouni, Marco Mellia, Marco Ajmone Marsan, Stefan Alfredsson, Jonas Karlsson, Anna Brunstrom, Ozgu Alay, Andra Lutu, Cise Midoglu, and Vincenzo Mancuso. 2017. Speedtest-like Measurements in 3G/4G Networks: The MONROE Experience. In *Teletraffic Congress (ITC 29), 2017 29th International*, Vol. 1. IEEE, 169–177.
[18] Ali Safari Khatouni, Martino Trevisan, Leonardo Regano, and Alessio Viticchié. 2017. Privacy issues of isps in the modern web. In *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 588–594.
[19] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, Jeff Bailey, Jeremy Dorfman, Jim Roskind, Joanna Kulik, Patrik Westin, Raman Tenneti, Robbie Shade, Ryan Hamilton, Victor Vasiliev, Wan-Teh Chang, and Zhongyi Shi. 2017. The QUIC Transport Protocol: Design and Internet-Scale Deployment. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication* (Los Angeles, CA, USA) *(SIGCOMM '17)*. ACM, New York, NY, USA, 183–196. https://doi.org/10.1145/3098822.3098842
[20] Vincenzo Mancuso and et al. 2019. Results from running an experiment as a service platform for mobile broadband networks in Europe. *Computer Communications* 133 (2019), 89 – 101.
[21] Anna Maria Mandalari, Andra Lutu, Ana Custura, Ali Safari Khatouni, Özgü Alay, Marcelo Bagnulo, Vaibhav Bajpai, Anna Brunstrom, Jörg Ott, Marco Mellia, and Gorry Fairhurst. 2018. Experience: Implications of Roaming in Europe. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (New Delhi, India) *(MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 179–189. https://doi.org/10.1145/3241539.3241577
[22] Cise Midoglu. 2020. Large Scale Speedtest Experimentation in Mobile Broadband Networks. https://mosaic-simulamet.com/large-scale-speedtest-experimentation-in-mobile-broadband-networks/. [Online; accessed 24-Jul-2020].
[23] Cise Midoglu, Leonhard Wimmer, Andra Lutu, Özgü Alay, and Carsten Griwodz. 2018. MONROE-Nettest: A configurable tool for dissecting speed measurements in mobile broadband networks. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 342–347.
[24] Miguel Peón-Quirós, Vincenzo Mancuso, Vincenzo Comite, Andra Lutu, Özgü Alay, Stefan Alfredsson, Jonas Karlsson, Anna Brunstrom, Marco Mellia, Ali Safari Khatouni, and Thomas Hirsch. 2017. Results from Running an Experiment as a Service Platform for Mobile Networks. In *Proceedings of the 11th Workshop on Wireless Network Testbeds, Experimental Evaluation & CHaracterization* (Snowbird, Utah, USA) *(WiNTECH '17)*. Association for Computing Machinery, New York, NY, USA, 9–16. https://doi.org/10.1145/3131473.3131485
[25] Mohammad Rajiullah, Andra Lutu, Ali Safari Khatouni, Mah-Rukh Fida, Marco Mellia, Anna Brunstrom, Ozgu Alay, Stefan Alfredsson, and Vincenzo Mancuso. 2019. Web Experience in Mobile Networks: Lessons from Two Million Page Visits. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1532–1543. https://doi.org/10.1145/3308558.3313606
[26] Jingjing Ren, Daniel J. Dubois, David Choffnes, Anna Maria Mandalari, Roman Kolcun, and Hamed Haddadi. 2019. Information Exposure for Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. In *Proc. of the Internet Measurement Conference (IMC)*.
[27] Ali Safari Khatouni, Francesca Soro, and Danilo Giordano. 2019. A Machine Learning Application for Latency Prediction in Operational 4G Networks. In *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. 71–74.
[28] Arunan Sivanathan, Hassan Habibi Gharakheili, Franco Loi, Adam Radford, Chamith Wijenayake, Arun Vishwanath, and Vijay Sivaraman. 2018. Classifying IoT devices in smart environments using network traffic characteristics. *IEEE Transactions on Mobile Computing* 18, 8 (2018), 1745–1759.
[29] Srikanth Sundaresan, Walter Donato, Nick Feamster, Renata Teixeira, Sam Crawford, and Antonio Pescape. 2012. Measuring Home Broadband Performance. *Commun. ACM* 55 (11 2012), 100–109. https://doi.org/10.1145/2366316.2366337
[30] Martino Trevisan, Idilio Drago, Marco Mellia, and Maurizio M. Munafò. 2016. Towards Web Service Classification using Addresses and DNS. In *Proceedings of the 7th International Workshop on Traffic Analysis and Characterization (TRAC'16)*. 38–43. https://ieeexplore.ieee.org/document/7577030/
[31] 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.
[32] Martino Trevisan, Danilo Giordano, Idilio Drago, Maurizio 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.
[33] Martino Trevisan, Ali Safari Khatouni, and Danilo Giordano. 2020. ERRANT: Realistic emulation of radio access networks. *Computer Networks* 176 (2020), 107289. https://doi.org/10.1016/j.comnet.2020.107289