Measuring Roaming in Europe: Infrastructure and Implications on Users QoE

*Original*
Measuring Roaming in Europe: Infrastructure and Implications on Users QoE / Lutu, Andra; Trevisan, Martino; Safari Khatouni, Ali; Mandalari, Anna Maria; Custura, Ana; Mellia, Marco; Alay, Ozgu; Bagnulo, Marcelo; Bajpai, Vaibhav; Brunstrom, Anna; Ott, Jorg; Fairhurst, Godred. - In: IEEE TRANSACTIONS ON MOBILE COMPUTING. - ISSN 1536-1233. - ELETTRONICO. - 21:10(2022), pp. 3687-3699. [10.1109/TMC.2021.3058787]

*Publisher:*
IEEE

*Published*
DOI:10.1109/TMC.2021.3058787

# Measuring Roaming in Europe: Infrastructure and Implications on Users' QoE

Andra Lutu[1], Martino Trevisan[2], Ali Safari Khatouni[3], Anna Maria Mandalari[4], Ana Custura[5],
Marco Mellia[2],Özgü Alay[6,7], Marcelo Bagnulo[8], Anna Brunstrom[9],
Vaibhav Bajpai[10], Jörg Ott[10], Gorry Fairhurst[5]
Telefonica Research[1] Politecnico di Torino[2] Western University[3] Imperial College London[4] University of Aberdeen[5] Simula Metropolitan[6] University of Oslo[7] Universidad Carlos III de Madrid[8] Karlstad Universitet[9] Technische Universität München[10]

**Abstract**—"Roam like Home" is the initiative of the European Commission to end the levy of extra charges when roaming within the European region. As a result, people can use data services more freely across Europe. However, the implications of roaming solutions on network performance have not been carefully examined yet. This paper provides an in-depth characterization of the implications of international data roaming within Europe. We build a unique roaming measurement platform using 16 different mobile networks deployed in 6 countries across Europe. Using this platform, we measure different aspects of international roaming in 4G networks in Europe, including mobile network configuration, performance characteristics, and quality of experience. We find that operators adopt a common approach to implement roaming called Home-routed roaming. This results in additional latency penalties of $60$ ms or more, depending on geographical distance. This leads to worse browsing performance, with an increase in the metrics related to Quality of Experience (QoE) of users (Page Load time and Speed Index) in the order of 15-20%. We further analyze the impact of latency on QoE metrics in isolation and find that the penalty imposed by Home Routing leads to degradation on QoE metrics up to 150% in case of intercontinental roaming. We make our dataset public to allow reproducing the results.

◆

## 1 INTRODUCTION

International roaming allows mobile users to use their voice and data services when they are abroad. The European Commission (EC), in an effort to create a single digital market across the European Union (EU), has (as of June 2017) launched a set of regulations [1] as part of the "Roam like Home" initiative. It abolishes additional charges for users when they use voice and data services while roaming in the EU. It is designed to prevent unexpected charges due to extra mediation and billing costs when roaming services are active. In this setting, Mobile Network Operator (MNO) are expected to deliver services with Quality of Service (QoS) properties similar to the ones a user experiences when at home.

In this paper, we present an extensive large-scale measurement study to understand the roaming ecosystem in Europe after the "Roam like Home" initiative. More specifically, we investigate: (i) Which technical solutions are actually being deployed and used today? and (ii) What are the implications of roaming on the service experienced by the roaming user? This paper extends our previous work [25] with new measurements and a thorough analysis of the implications of roaming on web browsing performance, that we use as an example of the impact of technical choices.

To address these questions, we build on the top of a unique measurement platform, *MONROE-Roaming*, to assess roaming and its performance implications. The platform integrates dedicated measurement hardware that we deployed in six different countries across Europe, covering a total of 16 MNOs. We purchased regular Subscriber Identity Modules (SIMs) that support roaming for these MNOs and distributed them across the six countries. Using this setup, we characterize roaming operation and network performance and evaluate the impact on web applications while roaming. Our experimental campaign includes a wide range of measurements, including `traceroute`, DNS lookups, HTTP downloads and fully automated webpage visits.

We find that all observed MNOs use home-routed roaming (HR), meaning that all the user's traffic is routed through the home network. This has several implications. First, HR translates in a latency penalty when roaming due to the longer paths that packets travel. For instance, when visiting a Spanish website when in Spain with a Norwegian SIM, packets have to travel between Spain and Norway. Second, all services will be available in the same way as in the home network, and the use of the home network DNS server implies that users would be redirected to CDN content at their home network. Third, we do not observe traffic differentiation policies for VoIP or web. Yet, content-based filtering and geo-blocking policies complicate the picture, as home-country rules would apply. Forth, we observe that the additional latency imposed to roaming users may negatively affect web browsing Quality of Experience (QoE). Indeed, our measurements show that popular QoE-related metrics (i.e., *onLoad* and *SpeedIndex*) increase in the order of 15-20% when roaming. Additional controlled experiments confirm the negative effect of network latency on web browsing performance, showing that intercontinental roaming can degrade the metrics up to 150%. While we were not able to check given the limitations of our platform, we agree that the same delay would impact applications too, since HR applies to all traffic. We release our dataset to stimulate further analyses and allow reproducing our

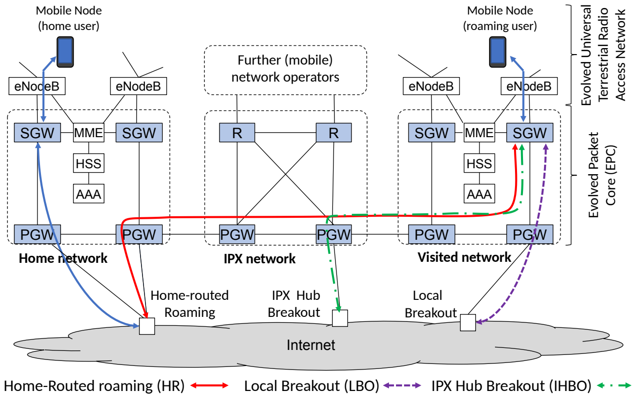


Figure 1: Internet access options for a mobile node at home (left) and when roaming (right).

results. It is available online, along with access instructions and a description of the tables.[1]

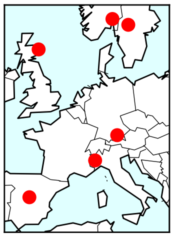
The remainder of the paper is organized as follows. Section 3 describes the MONROE-Roaming platform, while Section 4 illustrates the collected measurements. In Section 5, we describe our findings in terms of roaming configurations and their implications, while Section 6 focuses on the QoE of web users. In Section 7, we further quantify the implications of "Roam like Home" on users' QoE through controlled experiments. Sections 8 discusses our results and related work. Finally, Section 9 concludes the paper.

## 2 BACKGROUND

To support roaming, MNOs commonly connect with each other through an IP Packet Exchange (IPX) network. An IPX [3], [2] can be described as a hub that interconnects MNOs over a private IP backbone network and is possibly run by a third party IPX provider. An IPX provider has connections to multiple network operators and thus enables each MNO to connect to other operators via a single point of contact. The interconnections between MNOs are accompanied by roaming agreements that enable the operators to apply policies, control network access for roaming subscribers, and manage their roaming services. In Figure 1, we present a set of topology architectures that can be used for roaming in a mobile network, namely, HR (solid red path), local breakout (LBO) (dashed purple path), and IPX hub breakout (IHBO) (dotted green path).

With **HR** [4], [5], the IP address of the roaming user is provided by the home network. All traffic to and from the mobile user is routed through the home network, for which a GPRS Tunneling Protocol (GTP) tunnel is set up between the Serving Gateway (SGW) of the visited network and the Packet Data Network Gateway (PGW) of the home network. With the IP endpoint in the home network, all services will be available in the same way as in the home network.

When **LBO** [4], [5] is in place, the IP address of the roaming user is provided by the visited network. The GTP tunnel is terminated at the PGW of the visited network and



| | MNOs | | |
|---|---|---|---|
| NO | Telia NO | Telenor NO | |
| SE | Telia SE | Telenor SE | 3 SE |
| UK | Vodafone UK | EE | |
| DE | Vodafone DE | T-Mobile | O2 |
| ES | Vodafone ES | Movistar | Orange |
| IT | Vodafone IT | TIM | 3 IT |

Figure 2: The distribution (left) of the MONROE-Roaming nodes in six countries and (right) SIMs for 16 MNOs we measure across Europe. Each country deploys two MONROE-Roaming nodes and one measurement server.

IP-based services can be accessed directly from there. This does not add latency and reduces network resource usage, but may restrict access to private services in the user's home network. Service control and charging also become more complex using LBO.

**IHBO** [6] provides an alternative to overcome the limitations of home-routed roaming and local breakout. Here, the IP address of the roaming user is provided by the IPX network. The GTP tunnel from the SGW in the visited network terminates at a PGW in the IPX network. There may be multiple PGWs so that latency and resource usage can be reduced by selecting one geographically close to the visited network. As the IPX network maintains a trusted relationship with the home network, it may assign an IP address recognized by the home network to the roaming user, thereby allowing the user access also to private services in the home network. IHBO can also simplify setup and management as a single GTP tunnel, terminated in the IPX network, can be used for roaming users from different home networks.

The topology can have a potential impact on communication performance. For instance, when the node accesses services inside the *visited network*, the performance is likely to be worse in the HR case, because all packets travel twice between the visited and the home country; less so when the communication peer is in a third country and shall be minimal when accessing services in the home country. This may also have implications in the selection of Content Delivery Network (CDN) when roaming abroad because the mobile user will access a server in the home network rather than one close to their location. At last, this has implications on content-based filtering and geo-blocking policies as home rules would apply. We seek to understand which of these routing schemes are currently in use and which are their QoE implications.

## 3 PLATFORM AND MEASUREMENT SETUP

In this section, we present the hardware platform we use for roaming measurements and how we orchestrate it to collect our data.

### 3.1 MONROE-Roaming Platform

We design and build MONROE-Roaming, a dedicated platform for roaming measurements in Europe. MONROE-Roaming integrates several components that we depict in

---

[1] https://smartdata.polito.it/measuring-roaming-in-europe-infrastructure-and-implications-on-users-qoe/

Table 1: Terminology. We take the perspective of a given mobile user that has a SIM card registered with a service provider in her home network.

| | |
|---|---|
| **Home SIM** | The mobile user SIM while in the home country and connected to the home network. |
| **Roaming SIM** | The mobile user SIM while she is roaming in a foreign country |
| **Visited SIM** | A SIM subscribed to visited network in the visited country. |
| **Home Network** | The network to which the mobile user subscribes. |
| **Visited Network** | The network to which a user connects while roaming internationally in the visited country. |
| **Home Server** | A server located in the home network of the user's SIM. |
| **Visited Server** | A server located in the network to which user roams, i.e., the visited network. |

Figure 3. The main blocks include measurement nodes distributed in six different EU countries, the backend system, several measurement servers, and a scheduler, all of which we detail next. To build the MONROE-Roaming platform we adapted the open-source software provided by MONROE [10], an open measurement platform. [2]

**MONROE-Roaming nodes:** Each MONROE-Roaming node is equipped with an APU board from PC Engines with two 3G/4G MC7455 LTE CAT6 miniPCI express modems. Because of the high cost of nodes and subscriptions, and the complexity of the coordination effort required (see Section 3.3), we have set up a platform with a total of 12 MONROE-Roaming nodes dedicated for roaming measurements.

**MONROE-Roaming backend:** Upon completion of each measurement, MONROE-Roaming nodes transfer the measurement results to the backend for further analysis.

**Measurement servers:** We have deployed one measurement server in each country as measurement responders and also to capture traffic traces.

**MONROE-Roaming scheduler:** The scheduler allows the user to query for resources, select nodes and launch different tests in the platform simultaneously. We used the open-source MONROE scheduler as a basis for the MONROE-Roaming scheduler. Each test is designed and implemented in a Docker container [28].

### 3.2 Experimental Setup

To understand the roaming ecosystem in Europe, we focus on diversity of the MNOs. In other words, we aim to cover a large number of SIMs rather than running measurements from a large number of vantage points. To this end, we deployed two MONROE-Roaming nodes in each of the six European countries to measure a total of 16 MNOs that operate their own network, as illustrated in Figure 2.

For each MNO, we bought six SIMs that support roaming in Europe and we distributed one SIM in each of the countries we cover. For example, in Germany, we bought six Vodafone DE SIMs that support roaming. We kept one Vodafone DE SIM as the *home SIM* in the home country (i.e., Germany). Then, we distributed five *roaming SIMs* from Vodafone DE to the other five countries (i.e, Sweden, Norway, UK, Italy and Spain). Each roaming SIM connects to (or *camps on*) a local roaming partner (or visited network) native to the visited country. For example, Vodafone DE in Germany is a roaming partner of Telenor SE in Sweden. Therefore, Telenor SE serves Vodafone DE's customers roaming in Sweden by allowing Vodafone DE users to camp on Telenor SE's network. For each roaming SIM, we identify the corresponding visited network (e.g., Telenor SE in Sweden for Vodafone DE) and, when available, activate the corresponding native SIM from the visited network (which we hereinafter denote by *visited SIM*). We illustrate this configuration in the experimental setup in Figure 3. We also describe the terminology in Table 1.

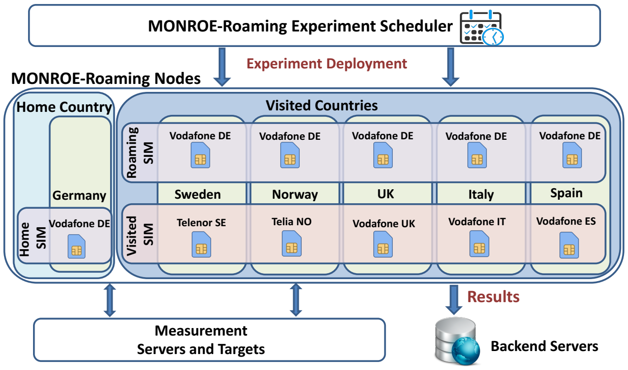


Figure 3: MONROE-Roaming platform and experimental setup. We exemplify our setup for Vodafone DE. We have five Vodafone DE SIMs in international roaming nodes and another SIM in the home country node. For each roaming Vodafone DE SIM, we insert the SIM corresponding to the local roaming partner for the MNO. For example, in Sweden, we use the Telenor SE SIM which corresponds to the network on which the Vodafone DE SIM is camping.

With our experimental setup, we can measure at the same time the Roaming SIM, the Home SIM and the Visited SIM. One SIM card was kept in the home country (where it was purchased), and the other five SIM cards were shipped to the visiting countries where these SIM cards will be roaming during the measurements. For each set of measurements, we evaluate the roaming service offers by a single MNO while roaming in different countries.

### 3.3 Measurement Coordination

Each MNO-specific measurement campaign involves 11 SIMs and 6 nodes: (i) one node with the home SIM and (ii) five nodes with both the roaming SIM and the corresponding visited SIM, as illustrated in Figure 3. This enables us to capture performance metrics for the roaming SIM, but also to compare those with the local performance of the home network and the visited network (when possible).

Before running the set of measurements (see Section 4), we first need to configure the nodes by activating and deploying the SIMs. For each MNO, we perform the measurements at the same time from all six countries and coordinate the configuration of the experimental setup in two steps:

**Home and Roaming User Activation:** To measure a MNO, we first insert the SIM into the first SIM slot in each node in all six deployment locations. For the SIM active in its home country, this step triggers the home user activation

---

[2] https://www.monroe-project.eu/access-monroe-platform/

(by inserting the SIM in the measurement node). For the rest of the nodes, this step triggers the roaming user activation.

**Visited User Activation:** Once we complete the home and roaming user activation, we check which visited network the roaming SIM uses in each of the five visited countries. Then, we insert the SIM of each partner MNO (when available) into the second slot of each corresponding node.

Using the MONROE-Roaming scheduler, we orchestrate the execution of the measurements so that they run in parallel on all nodes. The measurement coordination effort was a significant part of the process. In each country, at least one person was dedicated to carry out the physical experimental setup configuration for each MNO promptly. Given that we deploy two nodes per country, we could measure two MNOs and (maximum) 22 SIMs in parallel. We coordinated the SIM changes over email. Furthermore, before the change of the next pair of SIMs, we double-checked the measurement results we had collected to ensure correctness and completeness of the dataset. Each round lasted one week, over a total period of more than four months of experiments. We collect measurements continuously, without restricting to any specific time of the day.

# 4 MEASUREMENTS

## 4.1 Roaming Infrastructure Measurements

We run a series of measurements that enable us to identify the roaming setup, infer the network configuration for the 16 MNOs that we measure and quantify the end-user performance for the roaming configurations which we detect. We run `traceroute` for path discovery, `dig` for Domain Name Service (DNS) lookups and `curl` for testing HTTP data transfers with popular URLs. We complement this analysis with metadata (e.g., radio access technology, signal strength parameters) collected from each node. We do not run extensive *throughput* measurements, as the SIM data quotas make it difficult to run large-scale experiments. Moreover, throughput depends on many factors (TCP settings and cross-traffic among all), and, as such it would be difficult to draw solid conclusions. The reader can find in our previous work [22] a discussion on speedtest-like measurements on mobile networks and on online posts[3], which confirm our findings.

For each MNO, we measure in parallel the roaming user, the home user and the visited user (see Section 3 for terminology) through the MONROE-Roaming scheduler. In this way, we are able to capture potential performance penalties that might result, for example, from roaming internationally under a home-routed configuration. We performed measurements using both 3G and 4G networks to evaluate the impact of potentially different configurations for the two radio access technologies.

Next, we describe each measurement test and its resulting dataset in more detail.

**`traceroute`**: We run periodic `traceroute` measurements against all the servers we deploy in each country as measurement responders. We repeat the measurements ten times towards each target. The resulting dataset lists the set of IP hops along the forwarding paths from each vantage point towards each measurement responder. Additionally, we collect the public IP address for each vantage point (i.e., the IP endpoint associated with the mobile client as seen from the public Internet). We use `traceroute` for RTT measurements as well, since we collect and store the latency to the target and intermediate nodes as provided by the `traceroute` command-line tool.

**`dig`**: We run the `dig` utility for DNS lookups (over UDP port 53) against a list of 180 target Fully Qualified Domain Names (FQDNs) mapped to advertisement services. We use the independent filter lists from https://filterlists.com to build the list of targets. We focus on ad services because this type of third party services inflates significantly performance metrics of web services (e.g., page load time), as well as impact the web experience of mobile users [17]. Thus, it is important to capture (and potentially eliminate) any additional delay penalty that might impact how fast a roaming user receives this type of content. Each experiment uses the default DNS server for the tested MNO and queries for the A record associated with each of the target FQDNs. We store the entire output of each `dig` query, including the query time, the DNS server used and the A record retrieved. We repeat the `dig` queries twice for each FQDN from each vantage point, for a total of more than 2,000 queries per round.

**`curl`**: We run `curl` towards a set of 10 target popular webpages[4] over HTTP1.1/TLS. We repeat the measurements towards each URL at least 10 times (increasing the sample size if the SIM data quota allows it). We store various metrics, including the download speed, the size of the download, the total time of the test, the time to first byte, the name lookup time (query time) and the TCP handshake time.

**`metadata`**: We collect contextual information from the nodes, including the visited network Mobile Country Code (MCC) / Mobile Network Code (MNC) for each roaming SIM and the radio technology. This allows us to verify which visited network each roaming SIM uses as well as to identify and separate the collected data by radio technology.

## 4.2 Web Browsing Measurements

To deeper study the impact of roaming on web browsing performance, we run a specific measurement campaign. We leverage a customizable Docker container called MONROE-Browsertime [7] that we specifically engineered to run on any MONROE node, including the MONROE-Roaming nodes. We configured MONROE-Browsertime to mimic a mobile device browser (by setting both the screen resolution and the user-agent accordingly) to retrieve the mobile versions of the web pages. With it, we direct the browser to load a page and, at the end of page rendering, execute a custom Javascript script to collect a large number of metrics.

---

[3] https://www.speedtest.net/insights/blog/roaming-southeast-europe-2018/

[4] We target the following web pages: www.httpvshttps.com, facebook.com/telia/, en.wikipedia.org/wiki/Timeline_of_the_far_future, linkedin.com/company/facebook, www.yahoo.com/movies, instagram.com/leomessi/, google.com/search?q=iPhone+7,youtube.com/watch?v=xGJ5a7uIZ1g, ebay.com/globaldeals, nytimes.com, theguardian.com.uk/lifeandstyle.

We use the X virtual framebuffer (Xvfb) [9] for display emulation to let the browsers render the webpages. MONROE-Browsertime provides a configurable experiment template to enable web measurements in MONROE [32]. We configure each measurement by controlling (i) the network to test (the desired MBB interfaces, namely the roaming SIM and the corresponding visited SIM), (ii) the browser (in our case, we select Chrome 64.0.3282.186), and (iii) the web protocol (we run HTTP/1.1).[5] A combination of these parameters builds an *experiment setup*. We opted to use HTTP/1.1 to have a fair comparison among all websites since our goal is uniquely to spot performance variations under roaming. In our previous work [32], we run an extensive measurement campaign to assess the impact of HTTP/2 and QUIC under mobile networks, finding negligible differences in performance and users' QoE.

We select a list of 100 target pages to download from the SIMs of MNOs in Spain, Germany, Norway, and Sweden. We avoid the landing page in case it is too trivial (e.g., downloading `https://instagram.com/leomessi/` instead of `https://instagram.com/`). Our selection covers a wide range of user interests in terms of topics, including social networking, video, career, education, search engine, travel help, news, wiki, and shopping. Even more, we ensure that the list of target pages also contains web pages that are of interest mostly within the home country of the users (e.g., national news outlets). All websites expect TLS connections by default. We execute 10 different repetitions of measurements per target website. For each MNO, we measure in parallel the roaming user, the home user and the visited user where available (see Section 3 for terminology) through the MONROE-Roaming scheduler. We collect measurement from SIM of eight different operators (namely Orange ES, Vodafone ES, Telenor NO, Telia SE, Telenor SE, Tre SE, Telekom DE, and O2 DE) active in each of the countries that MONROE-Roaming covers (namely, ES, NO, SE, DE, IT and UK). In total, we perform 68 k visits to 100 URLs.

We track three main metrics that have been shown to be correlated with users' QoE [14]: Page Load Time (PLT), FirstPaint (FP) and RUMSpeedIndex (SI), which we detail next. The tool derives these metrics from browser timing metrics [27] that record the timing of different rendering phases of a page, from the initial DNS resolution to each HTTP request, from JavaScript processing to objects rendering.

**First Paint (FP):** It corresponds to the time when the browser starts rendering the first element of the page. This happens as soon as the first element of the page has been fetched and processed, and after the downloading of all needed elements (e.g., stylesheets).

**Page Load Time (PLT):** This is the time the last object on the page has been downloaded. It occurs when all the HTML files and any sub-resources (images, fonts, CSS, videos, etc.) are loaded. Note that not all these elements are needed to complete the rendering of the visible portion of the page.

**RUMSpeedIndex (SI):** It monitors the rendering process of the page by tracking the evolution of visible rectangles with elements that loaded external resources on a page. The original SpeedIndex requires to film the rendering process, and the postprocessing of the video to observe changes. Given the limited resources of the MONROE nodes (in terms of CPU, storage and communication bandwidth), we opt for the RUMSpeedIndex approximation, which uses the sequence of events as reported by the browser to estimate the time in which the visible portion of the screen would be modified [8]. Intuitively, it calculates the likely time that a paint event happens, given the processing and downloading of the various elements by the browser. The SI corresponds to the time when the last paint happens. This is considered a QoE approximation since it considers the evolution of the rendering process as seen by the user.

Given the interplay of objects, rendering, and visible portion of the page on the screen, there is no clear ordering on the metrics. For instance, the rendering of a page can start after one, or some, or all objects have been downloaded. Similarly, the rendering process may involve a portion of the area which is not currently on the visible part of the browser window. In addition, analytics objects are typically downloaded as last elements, after the rendering is completed, thus inflating the PLT. For this, it is consistent to compare results considering the same metric, but it is not appropriate to compare across different metrics.

Our platform does not allow us to test mobile apps. Given also the lack of well-accepted benchmarks to measure QoE on apps, we limit our tests to consider the mobile web browsing case. Albeit nowadays a large fraction of mobile traffic is generated by apps, still a significant fraction is due to web browsing. In addition, as apps commonly use HTTP as application protocol, we argue that mobile apps would suffer similar impairments of web applications when used in roaming.

## 5 ROAMING INFRASTRUCTURE

### 5.1 Roaming configuration

Our initial goal is to determine the roaming setup for each MNO (i.e., whether it used LBO, HR or IHBO). For this, we determine the MNO that allocates the public IP address of the roaming SIM. Our results show that *HR was used by all 16 MNOs from all the different roaming locations we capture*. We further corroborate this result by retrieving the first hop replying with a public IP address along the forwarding path from a roaming SIM to each server and identifying the MNO that owns it. We find that the first hop with a public IP address along the path lies in the original home network of each roaming SIM, which is consistent with HR. We repeated the experiments periodically, and always found that HR was still in place. Our last measurements were performed in June 2019.

### 5.2 Details on Roaming Infrastructure

We now evaluate the following performance metrics for each roaming SIM, home SIM and visited SIM: (i) the number of visited networks we observe for the roaming SIM, (ii) the number of hops from vantage point to target

---

[5] The nodes use the mobile carrier DNS resolver consistently with the operator currently in use.

measurement server, (iii) the number of home network PGWs that the roaming SIMs reached in comparison with the home network SIMs.

***Visited network selection***: The metadata we collect during the measurement campaign for each MNO enables us to verify the visited network that each roaming user camps on in the visited country. In general, we note stability both in 4G roaming and 3G roaming in the selection of the visited network (Table 2) in the six roaming locations. We also observe some differences between MNOs. For example, for Telekom DE, the 4G visited network chosen by each roaming SIM never changed during the measurement campaign, even when we forced the radio technology handover. This is consistent with all the six roaming locations. For O2 DE, on the other hand, the default 4G visited network did change over time for the SIMs roaming in Italy (3 visited networks), Norway (3 visited networks), and Sweden (2 visited networks). It should be noted that the length of the measurement period varies for each MNO, as it is impacted by multiple external factors (e.g., at times some of our measurement responders were affected by power outages or some SIM cards were not connecting to the 4G network due to poor coverage). This may influence part of the differences observed between the MNOs.

Table 2: Distribution of the first IP interface and visited network per MNO. We report the total number of networks each roaming user camps on in all visited countries (# of visited networks), the number of unique first IP addresses (# IP addr.), the total number of traceroutes we ran for the corresponding SIM (# tests) and the distribution for each first IP address we find (First hop breakdown(%)).

| MNO | Visited Networks | | IP addr. | Tests | First hop breakdown(%) |
|---|---|---|---|---|---|
| | 3G | 4G | | | |
| O2 DE | 9 | 9 | 20 | 657 | 1; 1; 1; 2; 2; 2; 2; 2; 3; 3; 3; 4; 4; 5; 6; 6; 9; 9; 11; 24 |
| Telekom DE | 5 | 5 | 4 | 1424 | 13; 19; 25; 43 |
| Voda DE | 5 | 6 | 2 | 1511 | 46; 54 |
| Movistar ES | 6 | 6 | 8 | 282 | 4; 5; 5; 7; 8; 21; 22; 28 |
| Orange ES | 7 | 7 | 3 | 900 | 6; 43; 51 |
| Voda ES | 5 | 5 | 1 | 1943 | 100 |
| TIM IT | 6 | 6 | 4 | 497 | 1; 1; 46; 52 |
| Voda IT | 5 | 5 | 4 | 759 | 19; 19; 23; 39 |
| Telenor NO | 5 | 5 | 3 | 398 | 8; 30; 62 |
| Telia NO | 5 | 5 | 4 | 379 | 7; 16; 38; 39 |
| 3 SE | 7 | 6 | 2 | 828 | 44; 56 |
| Telenor SE | 5 | 5 | 2 | 1362 | 32; 68 |
| Telia SE | 5 | 5 | 4 | 379 | 7; 16; 38; 39 |
| EE UK | 5 | 5 | 9 | 1038 | 3; 4; 4; 5; 8; 13; 17; 19; 27 |
| Voda UK | 5 | 5 | 1 | 503 | 100 |

***Traceroutes, number of IP hops***: We analyze our traceroute results from the roaming SIMs and compare them with the traceroute results we collect from the corresponding home SIM towards the same target server. For all MNOs we find that *the number of IP hops is the same*. [6] This is consistent with the HR configuration (Figure 1), where the GTP tunnel

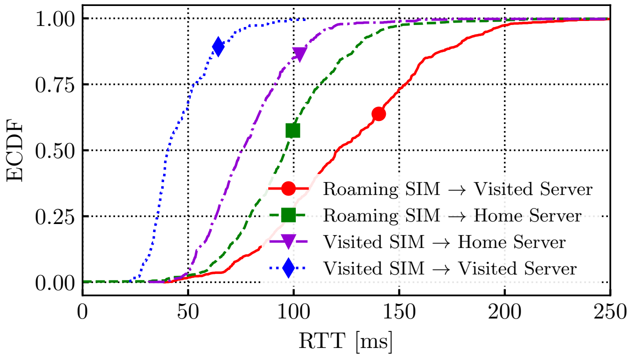


Figure 4: ECDF of the RTT from mobile nodes to target servers.

is defined between the SGW of the visited network and the PGW of the home network.

***Traceroutes, infrastructure***: By learning the IP addresses of the infrastructure elements along the forwarding path, we are able to infer aspects of the infrastructure deployment strategy of each MNO. In particular, by checking the IP address of the first hop in the path (Table 2), we find that MNOs have different strategies in terms of their deployments. We note that the first hops have an even distribution on their assignation to mobile users, showing that the MNOs have a similar approach for load balancing in their network. For example, for O2 DE we find 20 different first hops, suggesting that there might be a large number of PGWs deployed in the LTE infrastructure. For Vodafone UK, we see that the same first hop appears on the forwarding path, suggesting that the GTP tunnels of all our roaming users are terminated at a single PGW. We also note that although for the majority of MNOs these hops are configured with private address space [31], three operators (Telekom DE, Telenor NO, and Telenor SE) use public address space for their infrastructure. The last column in Table 2 details the breakdown of measurements among the number of different first hop IP addresses found. In some cases, a clear bias exists.

Finally, we verify that the set of first hops for roaming SIMs is the same as the set we observe from the home SIMs. This suggests that the roaming SIMs do not receive any differential treatment in terms of allocation to the PGWs. This is consistent for all MNOs we measure. Furthermore, when checking the 3G forwarding paths, we find that the set of IP addresses we see in 3G is a subset of the set of IP addresses we see in 4G, suggesting that the two functions are co-located in the same PGW [26]. We also check the time when the first IP address was used. We discover that all the PGWs are active at the same time. Multiple first IP addresses can be used at different times. We contacted 3 MNOs and the information they provided about their network confirms our findings.

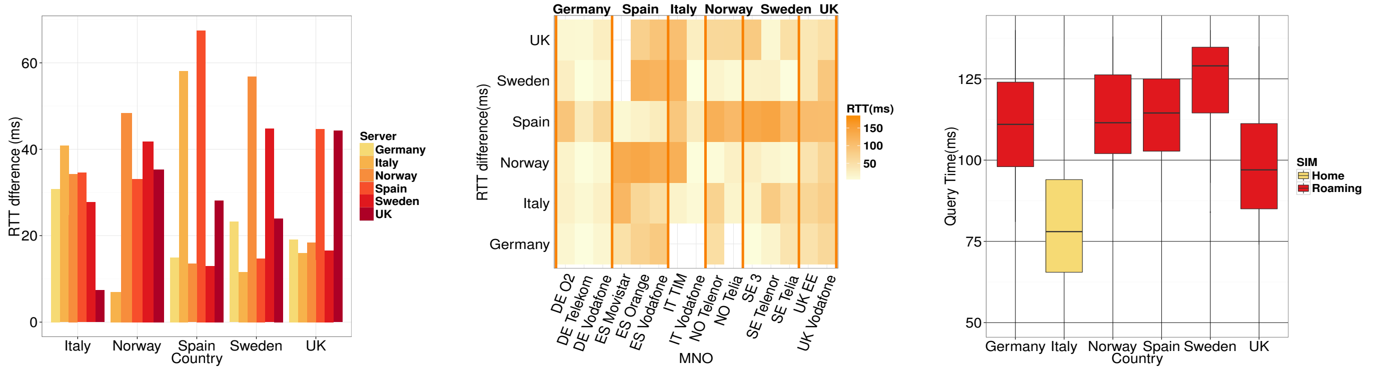
### 5.3 Implications of Home-Routed Roaming

***Delay implications***:

The HR data implies that the roaming user's exit point to the Internet is always in the original home country (Figure 1). Thus, the data always flows through the home

---

[6] Traceroute for 3 IT did not work in any country to any server.



(a) (b) (c)

Figure 5: Delay penalty of HR: (a) RTT difference from the visited country to all servers for Vodafone DE; (b) RTT difference per operator; (c) DNS Query time to all FQDNs for TIM IT.

network. Depending on the location of the server, this translates to a potential delay and performance penalty. Figure. 4 shows the ECDF of the RTT we measured between the roaming SIMs and the target servers located in the visited or home networks (red and green curves, respectively). We obtain the RTT values from the `traceroute` measurements using the latency measured toward the target. To compare the HR with the LBO configuration, we also include the RTT measurements between the visited SIMs against the same targets in the visited or home networks (blue and purple curves, respectively). The RTTs experienced by the visited SIMs serve as estimates of the RTTs that one could expect with a LBO configuration since LBO relies on access to local infrastructure with no need for tunneling back to the home network. We note that the largest delay penalty occurs when the roaming user tries to access a server located in the visited country. This is because the packets must go back and forth from the home network. Surprisingly, we note that the HR configuration also impacts the case when the roaming user accesses a target server located in the home network. That is, the GTP tunnel is slower than the native Internet path. In this case, the median value of the delay penalty considering all the MNOs is approximately 17ms. This varies across MNOs and in some cases we observe that the delay penalty is very low (e.g., just 0.2ms for O2 Germany). We check if the delay exhibits any daily periodicity, but found to evidence of it.

We investigate this performance impact further and calculate the estimated delay penalty between LBO and HR when the target is in the visited country. In more detail, we compute the delay penalty as the difference between the median delay to reach a given server when roaming, and the median delay to reach the same server from home. Figure 5a exemplifies these median values for Vodafone Germany. We note that, in general, the delay penalty varies widely with the geographical location of the roaming users and the target servers. For example, when a German SIM roams in Spain, the difference in terms of RTT is higher if the server is in the visited country (i.e., Spain) (red curve in Fig. 4). If the German SIM roams in Spain or Italy and the target server is in Norway or Sweden the delay penalty of the roaming is smaller, since to go to Norway or Sweden the data path would anyway likely pass through Germany.

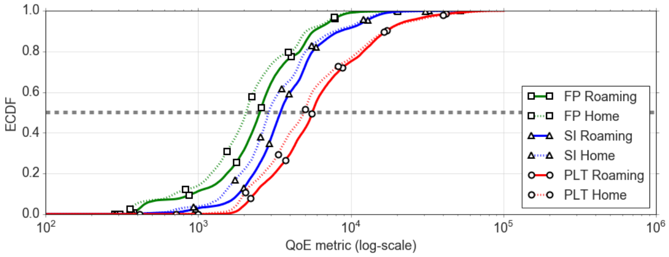
We then evaluate the difference between the roaming SIM RTT and the visited SIM TT towards the same target and we group them per MNO. Figure 5b shows the median value of the delay penalty of an MNO (on the `x` axis of the tile plot) while roaming against each of the six different servers (on the `y` axis of the tile plot, marked by country). We note that the delay penalty varies as a function of the location of the home country. For example, German SIMs experience a lower delay penalty, which is potentially due to them being in an advantageous position in the center of Europe.

***DNS implications***: The results of the `dig` measurements show that the DNS server offered to a roaming user is the same as the one offered when at home. This is again consistent with the use of HR. We verify whether this translates into an inflated query time for the roaming user. Figure 5c presents the distribution of DNS query times for all the SIMs of TIM IT. The query time is significantly lower for the home user than for the other five roaming users. This is consistent for all the 16 MNOs we measured. The usage of the home network DNS server further translates into implications in terms of CDN replica selection: the roaming user would be likely redirected to CDN content at its home network, and will not access the same content from a cache in the visited network (which would, in any case, result in facing a higher delay due to the home routing policy).

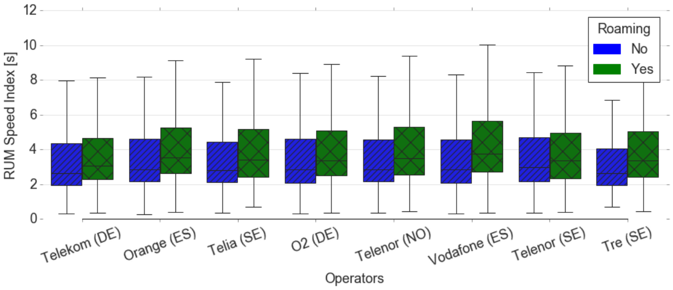
***HTTP performance implications***: Similar to the delay and DNS implications, international roaming affects HTTP and HTTPS performance. We quantify this penalty by considering the TCP handshake time between each SIM and the target web servers. The median value of the handshake time from the visited SIMs towards all the targets we measure is 170ms, while the median value for the roaming SIMs is 230ms. This leads to a delay penalty of approximately 60ms. As in the cases before, some MNOs are affected more by this roaming effect than others. In the following, we investigate the implications of web browsing in more detail.

## 6 WEB BROWSING WHILE ROAMING

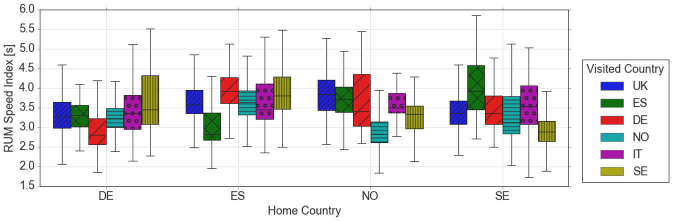
In this section, we report the results of our measurement campaign for web browsing. The results of our analysis show that, for any of the three QoE metrics we consider,



(a) Empirical CDF of QoE metrics (First Paint, RUM Speed Index and Page Load Time) while roaming in Europe and while in the home country..



(b) Boxplot of RUM Speed Index showing the impact of roaming broken down per home operator.

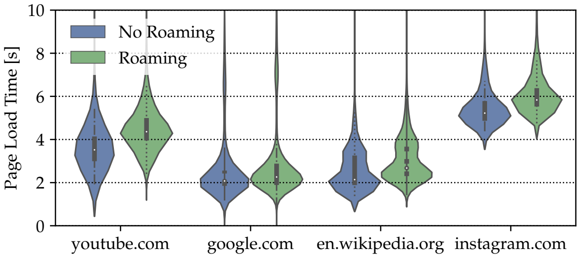


(c) Boxplot of RUM Speed Index showing the impact of roaming in specific visited countries (one box per visited country for each home country on the x-axis) compared to operating in the home country (on the x-axis).

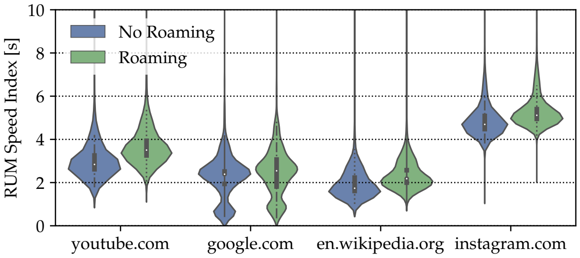
Figure 6: Impact of roaming on QoE metrics.

there is a penalty that roaming users suffer, which, in turn, may reflect into the users' perceived QoE. It is particularly interesting as poor QoE is known to damage the business model of big Internet players. Indeed, even small deterioration of quality levels could result in losses of revenues to providers.[7] Figure 6a shows the distribution of QoE metrics separately for roaming and non-roaming users. The curves do not overlap, as, for all websites, metrics have higher values. Indeed, median values for all increase in the order of $15-20\%$.[8]

We further break down the measurements by the home operator in Figure 6b. The boxplot corresponding to the roaming scenario aggregates all the measurements we collected in the countries where the corresponding SIMs from the operator (on the $x$ axis) were visiting. We observe that the roaming penalty is present for all operators to a similar extent. Figure 6c instead illustrates results separately by home and visited country. Similar considerations hold, with roaming users (where home and visited country differ) experiencing larger Speed Index.



(a) Page Load Time



(b) RUM Speed Index

Figure 7: Empirical QoE metric distributions with and without roaming for different target websites.

We then dissect a small set of target websites in Figure 7, namely YouTube, Google, Wikipedia, and Instagram. The figure shows with violin plots the distribution of Page Load Time (Figure 7a) and Speed Index (Figure 7b). The roaming penalty considerably varies depending on the target website. For instance, Google performance varies less than YouTube, with Wikipedia and Instagram lying in the middle. Moreover, YouTube Speed Index in roaming exceeds 5.8s in 10% of the cases, which is considered the threshold for a bad score.[9]

## 6.1 Finding most impacting factors

To capture how the web browsing QoE metrics correlate with the characteristics of the target website and with other browsing context features (e.g., radio coverage, roaming location), we generate Spearman's correlation matrix (Figure 8). This allows us to assess whether monotonic relationships (both linear or non-linear) exist between the QoE metrics and the features that we consider. We find that all QoE metrics correlate highly with the Time to First Byte (TTFB), the average number of images in the target webpages (images) and the average size of an image on the target webpage (image_size). The total number of objects (NoObj) and the total webpage size are also important, especially for the RUM Speed Index and the page load time.

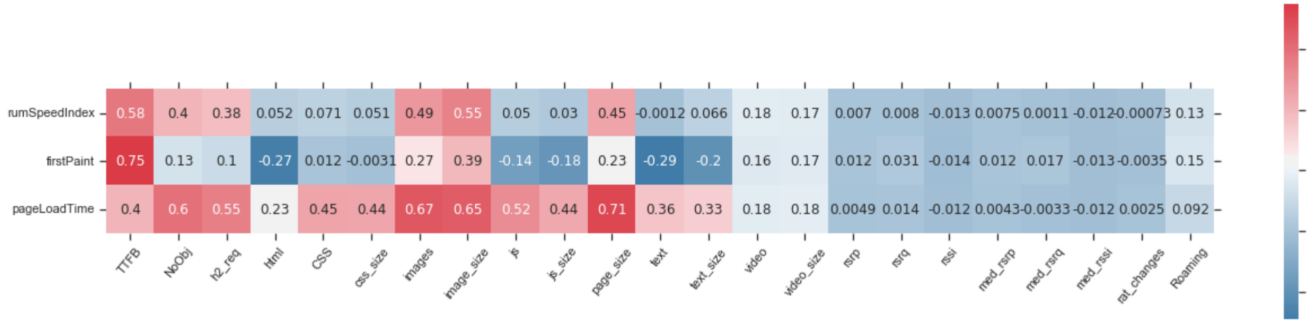
We focus on the RUM Speed Index QoE metric as the dependable variable and model it as a function of multiple predictors, including radio signal power, number of objects in the target website or the average size of images

---

[7] https://www.fastcompany.com/1825005/how-one-second-could-cost-amazon-16-billion-sales

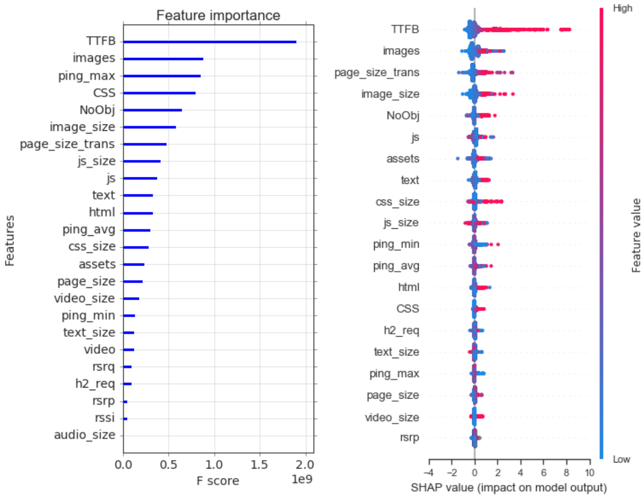
[8] Notice that we collected these measurements using HTTP/1.1 uniquely. We found HTTP/2 and QUIC having a negligible impact on users' QoE in our previous work [32].

[9] https://web.dev/speed-index/



| | TTFB | NoObj | h2_req | html | CSS | css_size | images | image_size | js | js_size | page_size | text | text_size | video | video_size | rsrp | rsrq | rssi | med_rsrp | med_rsrq | med_rssi | rat_changes | Roaming |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rumSpeedIndex | 0.58 | 0.4 | 0.38 | 0.052 | 0.071 | 0.051 | 0.49 | 0.55 | 0.05 | 0.03 | 0.45 | -0.0012 | 0.066 | 0.18 | 0.17 | 0.007 | 0.008 | -0.013 | 0.0075 | 0.0011 | -0.012 | -0.00073 | 0.13 |
| firstPaint | 0.75 | 0.13 | 0.1 | -0.27 | 0.012 | -0.0031 | 0.27 | 0.39 | -0.14 | -0.18 | 0.23 | -0.29 | -0.2 | 0.16 | 0.17 | 0.012 | 0.031 | -0.014 | 0.012 | 0.017 | -0.013 | -0.0035 | 0.15 |
| pageLoadTime | 0.4 | 0.6 | 0.55 | 0.23 | 0.45 | 0.44 | 0.67 | 0.65 | 0.52 | 0.44 | 0.71 | 0.36 | 0.33 | 0.18 | 0.18 | 0.0049 | 0.014 | -0.012 | 0.0043 | -0.0033 | -0.012 | 0.0025 | 0.092 |

Figure 8: Spearman’s correlation matrix between QoE metrics (rows) and browsing features (columns).



(a) Feature importance

(b) SHAP Value

Figure 9: Analysis showing the impact of different features on the RUM Speed Index. We show (a) The feature importance in the regression model as the average training loss reduction gained when using a feature for splitting; (b) the SHAP value variation showing the impact on the model output of the features we consider.

in the target webpage. We use the *xgboost* library [13] to train a regression model that can predict the SI from the set of features (i.e., predictors) we built. For the resulting prediction model, it is crucial to understand which features are the ones that impact the most on the end-user QoE. We aim for a model that is interpretable, where interpretability means that we can understand how the model uses input features to make predictions. We find that when analyzing the feature importance as the average training loss reduction gained when using a feature for splitting, the TTFB and the average number of images are the two features that bring the most information about the SI.

To further understand how different target website characteristics, together with browsing context features impact the QoE metrics, we use the SHapley Additive exPlanations (SHAP) value analysis [24]. SHAP is a game-theoretic approach to explain the output of any machine learning model. SHAP values represent a feature’s responsibility for a change in the model output. SHAP offers global interpretability — the collective SHAP values can show how much each predictor contributes, either positively or negatively, to the target variable, showing the positive or negative relationship for each variable with the target. Furthermore, SHAP also enables us to gain local interpretability — each observation gets its own set of SHAP values (see the individual SHAP value plot in Figure 9b). This greatly increases the model transparency. We can explain why a case receives its prediction and the contributions of the predictors. Traditional variable importance algorithms only show the results across the entire population but not on each individual case. The local interpretability enables us to pinpoint and contrast the impacts of the factors.

We investigate how the variations in the values of the different features impact the RUM Speed Index (see Figure 9). Similar to the prior observation, we note that the TTFB is an important feature whose value highly correlates with the QoE metrics of roamers: as the TTFB increases in value, so does the SHAP value (Figure 9b). This shows that the TTFB has a high impact on the roamers’ QoE metrics. Using the TTFB as a proxy for the location to which roamers travel, we conjecture that, given the prevalence of home routed roaming, the distance from the visited location to the roamers’ home country is an important predictor for the QoE metrics. We further investigate this dependency in Section 7.

We note that the five most important features that impact the web QoE (see Figure 9b) all stem from the structure of the target website and the location where the content is being served to the user. Thus, the number of images in the target webpage and javascript elements are among the most important factors that impact the roamer’s experience, together with the TTFB. Perhaps surprisingly, radio conditions that characterize the roamer’s mobile connection are far less important than the structure of the target website.

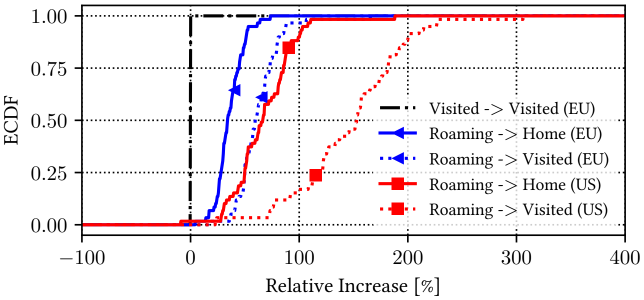
## 7 EMULATING DELAY TO MEASURE WEB BROWSING QOE

To further study the impact of the extra delay imposed to roaming users, we perform a large-scale measurement campaign in which we artificially vary network latency. The goal of these experiments is to study in isolation the impact of the latency on web browsing QoE, overcoming the limitations of the experiments with real roaming SIMs, in particular, obvious scalability issues. We automatically download web pages emulating home routing using different RTT curves and analyze variations in the classical QoE-related metrics.

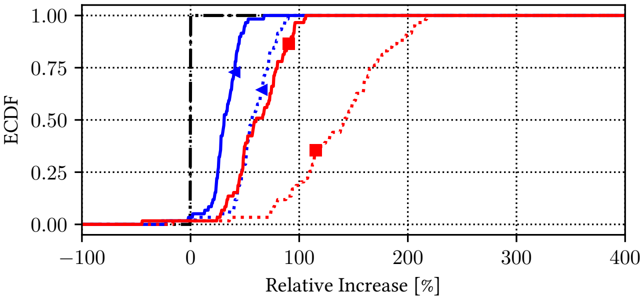
We deploy test machines in the six considered countries and configure proper traffic shaping rules to emulate the RTT observed in mobile experiments. To this end, we employ the `tc` Linux tool that allows sampling packet delay from a user-defined distribution. We use the RTT samples shown in Figure 4 for three scenarios: (i) *Visited -> Visited*, (ii) *Roaming -> Home* and (iii) *Roaming -> Visited*. We artificially create two additional RTT distributions, mimicking the scenario of a US user roaming in Europe. Given the impossibility of running real roaming experiments in the US, we employ the RIPE Atlas platform to collect ≈ 40 000 RTT samples between two servers located in the Netherlands and 680 RIPE Atlas nodes in the East coast of the US.[10] The resulting distribution is summed with samples of the *Visited -> Visited* distribution to add the typical mobile access network latency. We call the obtained distribution *Roaming -> Home (US)*, as it approximates the latency experienced by a US user roaming in Europe and accessing US websites. Median, 25th and 75th percentiles are respectively 160, 145 and 175 $ms$. We then create a second distribution simulating a US user roaming in Europe and accessing local European websites. In this case, packets travel from the EU to the US and back to the EU before reaching the server, whose response goes back to the US and the EU again. We emulate home-routed roaming by taking samples of EU to US RTT *twice*. We add access network latency as before. We call this distribution *Roaming -> Visited (US)*. Median, 25th and 75th percentiles are 265, 245 and 290 $ms$ respectively. Notice that in our previous work [35], we propose ERRANT, an open-source data-driven emulator which can enforce realistic traffic shaping profiles based on the MONROE measurements.

We instrument test machines to automatically download webpages using the *Browsertime* tool.[11] Given a URL, *Browsertime* starts the Chrome web browser, downloads the webpage, and collects statistics as described in Section 4.2. We build on SimilarWeb, a rank service analogous to Alexa to define the list of websites to test.[12] Separately by country, we consider the top-100 ranked websites for six categories (Food, Government, News, Shopping, Travel, and *all in all* rank), as well as 60 URLs of specific webpages of particular interest (social networks, search engines, etc.). In total, we download ≈ 2 700 webpages, as the lists partially overlap. Each test machine downloads the whole list, and with all the 5 RTT profiles. Each experiment is repeated 10 times, and in total, we perform 119 786 downloads over one week.

Once experiments are complete, we collect statistics from the test machines, and process data to compute QoE-related metrics for the different RTT profiles. For each webpage, we compute the median value when accessed as *Visited -> Visited* over the 10 repetitions. This represents our baseline, which emulates the condition of a non-roaming user accessing local websites. Then, we compute the median values for each RTT profile and compare them with the baseline. For each webpage $w$, we compute the relative increase of metric $m$ for profile $p$ from the *Visited -> Visited* profile as follows:



(a) Page Load Time



(b) RUM Speed Index

Figure 10: The relative deviation of objective metrics with different emulated RTT profiles. We observe worse performance with profiles with larger RTT.

$$increase(w, p, m) = \frac{median_{w,p}(m) - median_{w,v \to v}(m)}{median_{w,v \to v}(m)}$$

Figure 10 reports the results, separately for *Page Load Time* and *SpeedIndex*. Focusing on *Page Load Time* first (Figure 10a), we clearly note that, as RTT increases, webpages tend to be slower, with *Page Load Time* being, in median, 50% higher already with the *Roaming -> Home* profile. With other profiles the situation further worsens, with *Roaming -> Visited* 80% slower in median. The case of an EU citizen traveling the US and accessing local pages ( *roaming -> Visited (US)*) is noteworthy, here 170% slower than a non-roaming user. Very similar observations hold for *SpeedIndex* (Figure 10b), with a severe impact on performance of web browsing due to home-routing. Indeed, a Page Load Time larger than 3 seconds makes 53% of users to leave the website according to recent studies [11], while a Speed Index above 5.8 is considered too slow.[13] In this scenario, such large values are observed in 95% of cases.

In conclusion, our experiments study in isolation the impact of extra latency emulating the penalty imposed by HR in different scenarios. The results show a direct connection between network delay and slow page load time, which leads to a 150% impairment in the case of trans-continental roaming.

---

[10] https://atlas.ripe.net/

[11] https://www.sitespeed.io/documentation/browsertime/

[12] https://www.similarweb.com/

[13] https://web.dev/speed-index/

# 8 DISCUSSION

## 8.1 On Measurements Limitations

The MONROE-Roaming platform integrates measurement nodes located in six different EU countries and a measurement responder per country. Although this allows us to capture at a reasonable scale the performance of international roaming in Europe, it is still a limited view in terms of spatial sample distribution within each country (we only use two hardware devices per country, in the same location). Similarly, the findings in this paper represent one snapshot, calling for these experiments to be repeated over time to identify and investigate further changes. The high cost of deploying such an experimental study is a major restricting factor for the density of sampling geo-locations and prevented us from running *throughput* measurements, which would exhaust SIM card data quotas. We instead focus on characterizing multiple MNOs by taking advantage of the SIM farm we built using MONROE-Roaming. For each MNO, we purchased a similar data plan (10GB/month) enabling us to capture a similar number of samples per MNO and country. Furthermore, using the same equipment type throughout the measurement platform and in all locations eliminates any potential device bias we might observe in the measurement samples.

## 8.2 On Roaming Configurations

LBO appears a natural choice for an IP-based service and could offer lower operational costs as well as cheaper data tariffs. At the same time, we have shown that this can eliminate delay and potentially increase capacity for some traffic (depending on the destination). Although, LBO relies on access to local infrastructure, offering this could act as a product differentiator for the MNOs that provide this service first. In contrast, HR provides the home MNOs with all the accounting and billing information. This has been verified to be a major issue with MNOs that need to have a near real-time view of the customer traffic for accounting reasons.

Whereas Session Initiation Protocol (SIP) signaling could be used to derive billing information for voice (and VoLTE) calls, an MNO typically uses records to issue bills. Breakout at different points complicates this accounting, with possible abuse from customers (e.g, the delay in billing might allow an excessive amount of data traffic when roaming). Within the cellular network, classes of traffic can be differentiated using the Access Point Name (APN) and QoS Class Identifier (QCI). This could be used, for instance, by MNOs to implement HR for data, but LBO for VoLTE [20]. This raises the question of whether the roaming agreement could be updated using the same principle to break out some/all data traffic.

Any additional complexity from LBO can add to the operational cost of supporting users of the network (e.g., debugging issues, tracking faults, and predicting traffic). Further, if a service fails, it is not obvious who is responsible for finding the fault and fixing the issue. An IPX can help mitigate these impacts. Some solutions introduce additional proxy elects [2], responsible for routing traffic towards the correct network, and the associated control functions to coordinate.

Additionally, there are filtering rules, Digital Rights Management (DRM), language preference and personal content that depend upon the location (country) in which the content is viewed. The lawful intercept further complicates the picture. Here, the home network has full visibility of the necessary data, but the visited network may not. Lawful intercept may be further complicated because of variations in regulatory requirements depending on the geographic location of equipment. In a nutshell, enforcing and accounting for multiple policies for different content in different locations can become complex. Home routing simplifies this by letting the original operator monitor and manage all the traffic.

Lastly, access to content served by Content Delivery Networks (CDN) needs to be carefully optimized to avoid cases where a roaming user is redirected to a local replica that is spatially close, but whose network path is unnecessarily long (due to breakout constraints).

The choice of which form of roaming is used therefore is a function of the roaming agreement and capabilities of the visited network. These are constrained by many technical and legal requirements. Therefore, different breakout options can affect the performance of an application in different ways.

Because of these requirements, we do not expect the situation to change in the near future. Indeed, our most recent tests in June 2019 confirm that HR is still in place, despite all the performance implications.

## 8.3 Other implications

Traffic differentiation policies (such as blocking or throttling) may hamper Voice over IP (VoIP) communications for a roaming user in comparison to a home user. In our previous work [25], we ran a measurement campaign focusing on three popular VoIP applications: FaceTime, Facebook Messenger, and Whatsapp. We performed experiments making audio and video calls using each application. Results show that all operators allow users (even when roaming) to freely make VoIP calls using popular applications on their smartphones. Packet loss and bitrate are similar for roaming and non-roaming users. However, HR implies a non-negligible delay in communication, which is known to degrade users' perceived QoE. Indeed, a maximum latency of 150 ms is known to be the ideal for phone calls [33], [23]. This constraints would likely be violated when roaming internationally. For instance, making a VoIP call between two European mobile phones while in the US results in poor QoE due to excessive delay.

Another side effect of HR roaming is the complications of country-based content filtering. There are many reasons operators could have content filtering, which includes complying with government guidelines or following court orders, e.g., to restrict access to file-sharing websites in the UK [34], or the use of "opt-out" parental filters. In our previous work [25], we ran measurement campaigns using the software tests provided by the Open Observatory of Network Interference (OONI) [15] to detect censorship, surveillance and traffic manipulation. Results show that when censorship or blocking is present, it is the same in a home as a roaming case, consistently with HR. In other words, the content available in a user's home country remains available when roaming. As such, a roaming user is always subject to his home country rules, even when traveling in a foreign country where different laws are in place.

Because of data quota limitations, we were not able to run measurement campaigns for other data-intensive applications, such as video streaming. However, the latency penalty we observed in this analysis (for web browsing traffic) likely generalizes to these types of application traffic. Specifically, we conjecture that in the case of video streaming, we will observe a similar effect as we did for web browsing, where the content will be served to the users from a location close to their home country [16]. With the emerging immersive media formats and applications (e.g., 360-degree video, AR, VR), the latency budgets for the end-users diminishes even more, thus making HR roaming particularly problematic in the case of these applications.

### 8.4 Related Work

International roaming has received little coverage in terms of large measurement studies, potentially because of the high costs and coordination efforts associated with running such a campaign. Vallina *et al.* [36] have leveraged crowdsourced measurements and focused only on national roaming agreements between MNOs in France. The study does not provide further evaluation in terms of performance or content implications. Using controlled measurements in the dedicated platform MONROE [10] enabled Michelinakis *et al.* [29] analyze the impact of international roaming, but only for two operators in Europe. They find that the home-routed configuration does impact the performance of cloud service providers, such as Akamai or CloudFront. Our paper complements this work and presents an extensive measurement study to understand the international roaming ecosystem in Europe since the "Roam like Home" initiative.

We do not cover mobile measurement platforms and tools, but instead, refer the reader to existing surveys [18], [12] that cover them in great detail. There have been myriad recent studies focusing on mobile network characterization and performance. For instance, while Huang *et al.* [19] study LTE network characteristics in a cellular operator in the US, Safari *et al.* [22] show performance measurement in mobile networks are much more complex than wired networks, due to the different network configurations such as the presence of Network Address Translation (NAT) or Performance Enhancing Proxies (PEP), which do vary over time. Kaup *et al.* [21] run a crowdsourcing campaign to measure RTT and throughput towards popular websites in Germany. They used the dataset to show that the association of a mobile endpoint to the Point of Presence (PoP) within the operator network has an influence on network performance. The authors of [30] present a mobile app and a mechanism for identifying traffic differentiation for arbitrary applications in mobile networks. Ververis *et al.* [37] survey content filtering for a mixture of broadband and cellular ISPs and find a lack of transparency around the policies they implement as well as outdated and poorly implemented blacklists.

## 9 CONCLUSIONS

While roaming internationally, network configuration options can affect the performance of various applications for the end user. In practice, although there are three possible solutions (i.e., HR, LBO or IHBO), we find that HR is the norm for the MNOs we measured. This comes with performance penalties on the roaming user, who experiences increased delay and appears to the public Internet as being connected in the home country. This has further implications in the selection of CDN server replica when roaming abroad because the mobile user will access a server in the home network rather than one close to their location. We quantify the implications of HR on users' QoE with a specific measurement campaign and find that QoE-related metrics degrade in the order of 15-20% with respect to non-roaming users. Additional controlled experiments show that metrics degrade up to 150% in case of intercontinental roaming. HR has some immediate benefit to an operator. It simplifies billing and operational support and has an advantage that it provides a simple way to ensure the roaming user has consistent access (in the majority of cases) to CDN replicas and geo-restricted services from the home country in her native language.

We put these results in perspective while trying to also speculate on the commercial implications of the "Roam like Home" initiative. As regulation reduces the ability of MNOs to compete on price, the subscribers' quality of experience will potentially become a key factor in choosing a provider. We expect subscribers to increasingly compare the roaming experience to the home experience. Thus, an expectation of high quality, always-on services in a visited network follows and, if a home network fails to deliver in the visited network, the risk of churn increases. To this end, LBO is a natural step for an IP-based service and could offer lower operational cost, and cheaper tariffs for data, while at the same time we have shown this can eliminate delay and potentially increase capacity for some traffic (dependent on the destination). Although LBO relies on access to the infrastructure of the visited network, which can have implications on service control and charging, offering this could act in the advantage of the first operators to provide the service. Furthermore, in some cases, under the "Roam like Home" paradigm, some users may purchase SIMs from abroad to use in their country under permanent roaming conditions. Furthermore, the recent availability of commercial 5G service brings to our attention the question of 5G roaming Though we are not aware of any commercial 5G roaming available at the time of writing, we aim to explore this space in our future work.

## ACKNOWLEDGEMENTS

The research leading to these results has been funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 644399 (MONROE) and No. 777137 (5G-RANGE) and by the SmartData@PoliTO center for Big Data technologies. The work of Andra Lutu was supported by the EC H2020 Marie Curie Individual Fellowship 841315 (DICE).

## REFERENCES

[1] European Commission: New Rules on Roaming Charges and Open Internet. https://ec.europa.eu/digital-single-market/en/news/new-rules-roaming-charges-and-open-internet. [Online; 27-November-2019].

[2] GSM Association: Guidelines for IPX Provider Networks. https://www.gsma.com/newsroom/wp-content/uploads//IR.34-v13.0-1.pdf. [Online; accessed 27-November-2019].

[3] GSM Association: IPX White Paper. https://www.gsma.com/iot/wp-content/uploads/2012/03/ipxwp12.pdf. [Online; accessed 27-November-2019].

[4] GSM Association: LTE and EPC Roaming Guidelines. https://www.gsma.com/newsroom/wp-content/uploads/IR.88-v15.0.pdf. [Online; accessed 06-March-2018].

[5] Huawei: LTE International Roaming Whitepaper. http://carrier.huawei.com/en/technical-topics/core-network/LTE-roaming-whitepaper. [Online; accessed 06-March-2018].

[6] Method and System For Hub Breakout Roaming. https://patents.google.com/patent/US20140169286/en. [Online; accessed 06-March-2018].

[7] MONROE Browsertime. https://github.com/mrajiullah/browsertime-monroe. [Online; accessed 06-March-2018].

[8] RUM-SpeedIndex. https://github.com/WPO-Foundation/RUM-SpeedIndex. [Online; accessed 12-September-2019].

[9] XVFB. https://www.x.org/releases/X11R7.6/doc/man/man1/Xvfb.1.xhtml. [Online; accessed 06-March-2018].

[10] Alay, Ö., Lutu, A., Quirós, M. P., Mancuso, V., Hirsch, T., Evensen, K., Hansen, A. F., Alfredsson, S., Karlsson, J., Brunstrom, A., Khatouni, A. S., Mellia, M., and Marsan, M. A. Experience: An Open Platform for Experimentation with Commercial Mobile Broadband Networks. MobiCom 2017, pp. 70–78. http://doi.acm.org/10.1145/3117811.3117812.

[11] An, D. Find out how you stack up to new industry benchmarks for mobile page speed. 2018.

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

[13] Chen, T., and Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (2016), pp. 785–794.

[14] da Hora, D. N., Asrese, A. S., Christophides, V., Teixeira, R., and Rossi, D. Narrowing the gap between qos metrics and web qoe using above-the-fold metrics. In *International Conference on Passive and Active Network Measurement* (2018), Springer, pp. 31–43.

[15] Filastò, A., and Appelbaum, J. OONI: Open Observatory of Network Interference. USENIX FOCI 2012. https://www.usenix.org/conference/foci12/workshop-program/presentation/filastò.

[16] Giordano, D., Traverso, S., Grimaudo, L., Mellia, M., Baralis, E., Tongaonkar, A., and Saha, S. Youlighter: A cognitive approach to unveil youtube cdn and changes. *IEEE Transactions on Cognitive Communications and Networking 1*, 2 (2015), 161–174.

[17] Goel, U., Steiner, M., Wittie, M. P., Flack, M., and Ludin, S. Measuring What is Not Ours: A Tale of $3^{rd}$ Party Performance. In *Passive and Active Measurement* (2017), M. A. Kaafar, S. Uhlig, and J. Amann, Eds., Springer International Publishing, pp. 142–155.

[18] Goel, U., Wittie, M. P., Claffy, K. C., and Le, A. Survey of End-to-End Mobile Network Measurement Testbeds, Tools, and Services. *IEEE Communications Surveys and Tutorials 18*, 1 (2016). https://doi.org/10.1109/COMST.2015.2485979.

[19] Huang, J., Qian, F., Guo, Y., Zhou, Y., Xu, Q., Mao, Z. M., Sen, S., and Spatscheck, O. An in-depth study of LTE: effect of network protocol and application behavior on performance. *ACM SIGCOMM Computer Communication Review 43*, 4 (2013), 363–374.

[20] Kaltsas, I. Make Yourself at Home : A Comparative Study of VoLTE Roaming Architectures. Master's thesis, KTH, School of Information and Communication Technology (ICT), 2017. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-200335.

[21] Kaup, F., Michelinakis, F., Bui, N., Widmer, J., Wac, K., and Hausheer, D. Assessing the Implications of Cellular Network Performance on Mobile Content Access. *IEEE Transactions on Network and Service Management 13*, 2 (2016), 168–180. https://doi.org/10.1109/TNSM.2016.2544402.

[22] Khatouni, A. S., Mellia, M., Marsan, M. A., Alfredsson, S., Karlsson, J., Brunstrom, A., Alay, O., Lutu, A., Midoglu, C., and Mancuso, V. Speedtest-like measurements in 3g/4g networks: The monroe experience. In *2017 29th International Teletraffic Congress (ITC 29)* (2017), vol. 1, IEEE, pp. 169–177.

[23] Lewis, C., and Pickavance, S. Implementing Quality of Service Over Cisco MPLS VPNs. *Introduction to QoS* (2006).

[24] Lundberg, S. M., and Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774.

[25] Mandalari, A. M., Lutu, A., Custura, A., Safari Khatouni, A., Alay, Ö., Bagnulo, M., Bajpai, V., Brunstrom, A., Ott, J., Mellia, M., et al. Experience: implications of roaming in europe. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (2018), ACM, pp. 179–189.

[26] Marquez, C., Gramaglia, M., Fiore, M., Banchs, A., Ziemlicki, C., and Smoreda, Z. Not All Apps Are Created Equal: Analysis of Spatiotemporal Heterogeneity in Nationwide Mobile Service Usage. CoNEXT 2017. http://doi.acm.org/10.1145/3143361.3143369.

[27] Meenan, P. How fast is your web site? *Queue 11*, 2 (Mar. 2013). http://doi.acm.org/10.1145/2436696.2446236.

[28] Merkel, D. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal 2014*, 239 (Mar. 2014). http://dl.acm.org/citation.cfm?id=2600239.2600241.

[29] Michelinakis, F., Doroud, H., Razaghpanah, A., Lutu, A., Vallina-Rodriguez, N., Gill, P., and Widmer, J. The Cloud that Runs the Mobile Internet: A Measurement Study of Mobile Cloud Services. In *Proc. IEEE INFOCOM* (Honolulu, HI, USA, April 2018).

[30] Molavi Kakhki, A., Razaghpanah, A., Li, A., Koo, H., Golani, R., Choffnes, D., Gill, P., and Mislove, A. Identifying traffic differentiation in mobile networks. In *Proceedings of the 2015 Internet Measurement Conference* (New York, NY, USA, 2015), IMC '15, ACM, pp. 239–251.

[31] Moskowitz, R., Karrenberg, D., Rekhter, Y., Lear, E., and de Groot, G. J. Address Allocation for Private Internets. RFC 1918, Feb. 1996.

[32] Rajiullah, M., Lutu, A., Safari Khatouni, A., Rukh, M., Brunstrom, A., Alay, O., Alfredsson, S., and Mancuso, V. Web experience in mobile networks: Lessons from two million page visits. In *Proc. ACM Web Performance (www)* (02 2019).

[33] Recommendation, I. 114, one-way transmission time. *International Telecommunication Union* (1996).

[34] Rosati, E. 2015: the year of blocking injunctions? *Journal of Intellectual Property Law & Practice 10*, 3 (2015), 147.

[35] Trevisan, M., Khatouni, A. S., and Giordano, D. Errant: Realistic emulation of radio access networks. *Computer Networks 176* (2020), 107289.

[36] Vallina-Rodriguez, N., Sundaresan, S., Kreibich, C., Weaver, N., and Paxson, V. Beyond the radio: Illuminating the higher layers of mobile networks. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (2015), ACM, pp. 375–387.

[37] Ververis, V., Kargiotakis, G., Filastò, A., Fabian, B., and Alexandros, A. Understanding internet censorship policy: The case of greece. In *5th USENIX Workshop on Free and Open Communications on the Internet (FOCI 15)* (Washington, D.C., 2015), USENIX Association.

## BIOGRAPHIES

**Anna Maria Mandalari** is a research associate in the Dyson School of Design Engineering at the Faculty of Engineering at Imperial College London. Over the last four years she was a METRICS Marie Curie Early Stage Researcher affiliated with the University Carlos III of Madrid (UC3M). She has given invited talks in various research institutes and meetings: RIPE meetings, NEC laboratories (Heidelberg, Germany), the Max-Planck Institute for Informatics (Saarbrucken, Germany) and IETF meetings.

**Andra Lutu** Andra Lutu is an Associate Researcher at Telefonica Research in Barcelona, Spain. Her main research interests lie in the areas of network measurements, interdomain routing and mobile networks. After receiving her PhD at UC3M and IMDEA Networks Institute, she worked as a Postdoc Fellow at Simula Research Laboratory, where she was a main contributor to the H2020 MONROE project, helping to build the first open European hardware infrastructure to perform measurements in operational mobile networks. As part of Telefonica Research, Andra is a recipient of an H2020 MSCA Individual Fellowship grant funding her work on Dynamic Interconnections for the Cellular Ecosystem (DICE).

**Ana Custura** has an industry background is in Software and Network engineering, and is currently active in the field of Internet research at the University of Aberdeen. She received an MSc in Software Engineering and Network Management from Robert Gordon University in 2014, and her interests include large-scale Internet measurements, protocol design, and privacy and digital freedom. Ana contributes to the open source community maintaining tools for network measurement for the Debian Project and the Tor Project.

**Ali Safari Khatouni** is a postdoctoral associate at the Faculty of Computer Science at Western University in Prof. Hanan Lutfiyya and Prof. Michael Bauer's research group. He is currently teaching Computer Science Fundamentals I and Data Structures and Algorithms using Python. He was a postdoctoral fellow at the Faculty of Computer Science at Dalhousie University in Prof. Nur Zincir-Heywood's research group. He received his B.S. degree in Software engineering from Urmia University, Iran, and M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering at Politecnico di Torino, Italy. His research interests are in the areas of network traffic analysis, traffic classification, machine learning, mobile broadband networks, and smart city. He participated in several European research projects (Mplane and MONROE).

**Özgü Alay** [M] received the B.Sc. and M.Sc. degrees in electrical and electronic engineering from Middle East Technical University, Turkey, and the Ph.D. degree in electrical and computer engineering from the Tandon School of Engineering, New York University, USA. She is currently an Associate Professor with the University of Oslo, Norway. She is also the Head of Mobile Systems and Analytics (MOSAIC) Department at Simula Metropolitan, Norway. Her research team focuses on the empirical characterization of mobile networks and, the design of novel protocols and applications for future mobile networks.

**Marcelo Bagnulo** received the Electrical Engineering degree and the Ph.D. in Telecommunications in 2005, from Universidad Carlos III de Madrid (UC3M), Spain. He holds a tenured associate professor position at UC3M since 2008. His research interests include Internet architecture and protocols, inter-domain routing and security. He has published more than 70 papers in the field of advanced communications in journals and congresses and he is the author of 20 RFCs in the Internet Engineering Task Force (IETF) including the Shim6 protocol for IPv6 multihoming and the NAT64/DNS64 tools suite for IPv6 transition. Dr. Bagnulo was a member of the Internet Architecture Board between 2009 and 2011.

**Vaibhav Bajpai** is a senior researcher at TUM, Germany. He received his PhD (2016) and Masters (2012) degrees in Computer Science from Jacobs University Bremen, Germany. He is the recipient of the best of CCR award (2019), ACM SIGCOMM best paper award (2018), and IEEE COMSOC award (2017) for the best dissertation in network and service management. He is interested in future Internet protocols, web and video content delivery, network operations and management, and reproducibility of scientific Internet research.

**Anna Brunstrom** received a B.Sc. in Computer Science and Mathematics from Pepperdine University, CA, in 1991, and a M.Sc. and Ph.D. in Computer Science from College of William & Mary, VA, in 1993 and 1996, respectively. She joined the Department of Computer Science at Karlstad University (KAU), Sweden, in 1996, where she is currently a Full Professor and Research Manager for the Distributed Systems and Communications Research Group. Her research interests include Internet architectures and protocols, techniques for low latency Internet communication, multipath communication and performance evaluation of mobile broadband systems including 5G. She is currently the KAU principal investigator within the EU H2020 project 5GENESIS. She is a co-chair of the RTP Media Congestion Avoidance Techniques (rmcat) working group within the IETF. She has authored/coauthored over 200 international journal and conference papers.

**Jörg Ott** holds the Chair of Connected Mobility in the Department of Informatics at the Technical University of Munich. He is also Adjunct Professor for Networking Technology at Aalto University. He received his diploma and doctoral (Dr.-Ing.) degree in computer science from TU Berlin (1991 and 1997, respectively), and his diploma in industrial engineering from TFH Berlin (1995). His research interests are in network architecture, (Internet) protocol design, and networked systems, with a focus on (mobile) decentralized services and cloudless applications.

**Martino Trevisan** received his PhD in 2019 from Politecnico di Torino, Italy. He is currently a postdoctoral researcher at the SmartData@Polito center in the same university. He has been collaborating in both Industry and European projects and spent six months in Telecom ParisTech, France working on High-Speed Traffic Monitoring during his M.Sc. He visited twice Cisco labs in San Jose in summer 2016 and 2017, as well as AT&T labs during fall 2018. His research interests are Big Data and Machine Learning applied to Network Measurements and Traffic Monitoring.

**Marco Mellia** (M'97–SM'08) is full professor at the Electronics and Telecommunications Department of Politecnico di Torino. In 2002 he visited the Sprint Advanced Technology Laboratories working at the IP Monitoring Project (IPMON). In 2011, 2012, 2013 he collaborated with Narus Inc, CA, working on traffic monitoring and cyber-security system design. In 2015 and 2016 he collaborated with Cisco Systems for the design of cloud monitoring platforms based on machine learning. He is now the coordinator of the SmartData@PoliTO center on data science and machine learning, involving more than 50 colleagues and PhD students. Prof. Mellia co-authored over 250 papers published in international journals and presented in leading conferences, all of them in the area of communication networks. He won the IRTF ANR Prize at IETF-88, and best paper award at IEEE P2P'12, ACM CoNEXT'13, IEEE ICDCS'15. He is Area Editor of ACM CCR, IEEE Transactions on Network and Service Management and Elsevier Computer Networks. His research interests are in area of traffic monitoring and big data analysis, with applications to traffic classification, management and security.

**Gorry Fairhurst** received his first degree in Applied Physics and Electronics, and a PhD in Communications Engineering from the University of Aberdeen, UK, where he is a Professor specializing in Internet Engineering. He has experience in practical testing and benchmarking, and development of new Internet algorithms and protocols. His current research focuses on large-scale Internet measurement, Internet Transport, Satellite Broadband and the intersection of transport security and network operations. He is an active participant in the specification and engineering of standards, where he contributes to the work of ETSI and the IETF. He currently chairs the IETF transport and services working group and has authored over 25 RFCs.