5

MONROE: Measuring Mobile Broadband Networks in Europe

Özgü Alay[1], Andra Lutu[1], Rafael García[2], Miguel Peón-Quirós[2], Vincenzo Mancuso[2], Thomas Hirsch[3], Tobias Dely[3], Jonas Werme[3], Kristian Evensen[3], Audun Hansen[3], Stefan Alfredsson[4], Jonas Karlsson[4], Anna Brunstrom[4], Ali Safari Khatouni[5], Marco Mellia[5], Marco Ajmone Marsan[2,5], Roberto Monno[6] and Hakon Lonsethagen[7]

[1]Simula Research Laboratory
[2]IMDEA Networks
[3]Celerway Communications
[4]Karlstad University
[5]Politecnico di Torino
[6]Nextworks
[7]Telenor Research

Abstract

Mobile broadband (MBB) networks (e.g., 3G/4G) underpin numerous vital operations of the society and are arguably becoming the most important piece of the communications infrastructure. Given the importance of MBB networks, there is a strong need for objective information about their performance, particularly, the quality experienced by the end user. Such information is valuable to operators, regulators and policy makers, consumers and society at large, businesses whose services depend on MBB networks, researchers and innovators. In this chapter, we introduce the MONROE[1] measurement platform:

[1]MONROE is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644399. For more information, please visit https://www.monroe-project.eu/

An open access, European-scale, and flexible hardware-based platform for measurements and custom experimentation on operational MBB networks with WiFi connectivity. The platform consists of mobile and stationary nodes that are flexible and powerful enough to run most measurement and experiments tasks, including demanding applications like adaptive video streaming. Access to such a platform enables accurate, realistic and meaningful assessment of the performance of MBB networks by continuously monitoring these networks via active testing (e.g., delay test, web performance test, download speed test) and context metadata collection (e.g., connection mode, signal strength parameters). The multihoming feature of MONROE allows for the comparison of different networks under similar conditions as well as the exploration of new ways of aggregating providers to increase performance and robustness. In this chapter, we showcase the monitoring capabilities of the platform by analyzing preliminary performance measurement results. Considering that MONROE is *open* to external users, we further discuss a representative set of measurements and experiments to highlight the potential use cases of the platform. We argue that mobile measurements over operational networks, hence platforms such as MONROE, are crucial not only for characterizing and improving the user experience for services that are running on the current 3G/4G infrastructure, but also for providing feedback on the design of upcoming 5G technologies.

5.1 Introduction

Wireless and mobile access to the Internet have revolutionized the way people interact and access information. Mobile broadband (MBB) networks have become the key infrastructure for people to stay connected everywhere they go and while on the move. According to Cisco's Global Mobile Data Traffic Forecast [1], in 2015 the number of mobile devices grew to a total of 7.9 billion, exceeding the world's population. Also, fourth generation (4G) traffic exceeded third generation (3G) traffic for the first time in 2015 [1].

The society's increased reliance on MBB networks has made provisioning ubiquitous coverage the highest priority target for mobile network operators, as well as focusing on performance and user quality of experience (QoE). MBB coverage and performance experienced by the end-users are of great importance to many stakeholders including mobile subscribers, regulators, governments and businesses whose services depend on MBB networks. This also motivates researchers and engineers to further enhance the capabilities of mobile networks, by designing new technologies to cater for plethora of new applications and services, growth in traffic volume and a wide variety of user devices. In this dynamic ecosystem, there is a strong need for both open objective data about the performance and reliability of different MBB operators, as well as open platforms for experimentation with operational MBB providers. On the one hand, objective performance data is essential for regulators to ensure transparency and the general quality level of the basic Internet access service [2], especially in light of an evolution of service offerings beyond the best-effort traffic mode, including a balanced approach to net neutrality. On the other hand, custom experimental approaches are key to forwarding our understanding and driving innovation in MBB networks.

Characterizing the performance of home and mobile broadband networks requires systematic end to end measurements. Several regulators have translated this need into ongoing nationwide efforts, for example, the FCC's Measuring Broadband America initiative [3] in the USA. Operators and independent agencies sometimes perform drive-by tests to identify coverage holes or performance problems. These tests are, however, expensive and do not scale well [4]. Another approach is to rely on end users to run performance tests by visiting a website (e.g., [5]) or running a special measurement application (e.g., [6]). The main advantage of this approach is scalability: it can collect millions of measurements from different regions, networks and user equipment. However, with such an approach, repeatability is hard and one can only collect measurement data at users' own will, with no possibility to either monitor or control the measurement process. Furthermore, mostly due to privacy reasons, these measurements do not provide rich context information and metadata, e.g., location, type of user equipment, type of subscription, and connection mode (2G/3G/4G); however, metadata is critical when analyzing the results. Also, such a setup does not provide active measurements that can reveal important information on stability and availability of a network, since this requires long and uninterrupted measurement sessions. Finally, this approach limits the possibility of testing novel applications and services since this might require configuration changes (e.g., customized kernels).

MONROE is the first European platform for open, independent, multihomed, large-scale monitoring and assessment of performance of mobile broadband networks in heterogeneous environments. Access to such a platform allows for the deployment of extensive measurement campaigns to collect data from operational MBB networks. The availability of this vast amount of data allows us to advance our understanding of the fundamental characteristics of MBB networks and their relationship with the performance parameters of popular applications. This is crucial not only for improving the user experience for services that are running on the current 3G/4G infrastructure, but also for providing feedback on the design of upcoming 5G technologies.

In the remainder of this chapter, we summarize the current state of the art in Section 5.2. We then expand on the MONROE vision in Section 5.3, where we provide an overview of the MONROE goals and the key features of the measurement platform. In Section 5.4, we describe the current architecture design of the MONROE platform. We discuss in Section 5.5 how the MONROE user access and scheduling system is designed and how users can deploy their experiments. In Section 5.6, we present initial results from basic measurements running on operational MONROE nodes active in Norway, Sweden and Spain. We show that the MONROE system enables efficient MBB performance monitoring, operator benchmarking and complex network analytics. Finally, we conclude the chapter in Section 5.8.

5.2 Background and State of the Art

During the past years, we have seen increased interest in the networking community from different parties (e.g., researchers, operators, regulators, policy makers) in measuring the performance of mobile broadband networks. In this section, we aim to provide a condensed but comprehensive review of some of the most relevant approaches that strive to shed light on the mobile broadband ecosystem.

Large scale research measurement platforms such as RIPE Atlas [7], BISmark [8] or PlanetLab [9] share many common goals with MONROE. However, these platforms do not operate in the mobile environment. In order to cater to the need of open large-scale MBB measurements and to address the scarcity of available measurement platforms, several crowdsourcing approaches emerged over the past years, either from the research environment, e.g., Netalyzr [6], NetPiculet [10], or commercial-oriented, e.g., OpenSignal [11], RootMetrics [12] or MobiPerf [13]. These approaches leverage the wide adoption of mobile devices in the world and depend on the willingness of end-users to run the proposed tests. We note that the common vision of these tools is to identify and monitor a set of significant metrics which can accurately describe mobile broadband performance to the interested parties. For example, commercial-oriented OpenSignal proposes a complete approach for building MBB coverage maps by retrieving the connectivity-related metadata from user devices and characterizing multiple radio access technologies in the same ares. They introduce the notion of "time coverage" which provide s statistics for the time a device has been using a certain radio access technology in order to provide the end-user the possibility to make informed decisions in terms of the preferred MBB provider in a certain area. Similarly, RootMetrics defines a set of key performance metrics which allows for network benchmarking, with the intent of rating different providers available in a certain geographical area. Additionally, tools such as NetPiculet or Netalyzr aim to shed light on the infrastructure and the performance of broadband providers with the purpose of informing protocol and application design.

There are several research projects [6, 14–17] that use custom-designed apps to crowdsource and measure the performance of MBB providers and popular Internet applications, with a main focus on web browsing [18] and video streaming [19]. For example, MobiPerf [13] enables mobile network performance analysis [14]. The app builds on top of the Mobilyzer open library [20] and tracks a series of network performance metrics, such as HTTP benchmark downloading latency and bandwidth, traceroute with latency to different hops, ping latency, DNS lookup latency, TCP uplink and downlink throughput or RRC states metrics. Other similar relevant measurement efforts from the research community include [21–23].

With the increasing popularity of web and video-related services over MBB networks [24], there is a magnitude of research studies that focus on understanding the correlation between the network quality of service (QoS) metrics and the quality of experience (QoE) of the end-users [24–26]. In particular, this is appealing to operators, who continuously strive to provide the best service to their subscribers in order to increase their customer base. At the same time, the end-users themselves are looking for relevant metrics that can objectively assess the performance of popular applications over different MBB providers. In addition to the application performance, another important concern for the users is the energy efficiency of bandwidth intensive applications [27, 28].

Even more, alongside the attention coming from end-users, businesses or operators, there is rising interest from regulators for defining and monitoring a representative and unitary set of metrics that accurately captures the performance of today's broadband services in practice. In this sense, several of them (e.g., FCC, Ofcom and Anatel) have translated these efforts into national projects in collaboration with commercial partners such as SamKnows [29], which specializes in home and mobile broadband performance evaluation.

However, in order to allow for an open an unitary approach as well as the comparability of measurements, a common open framework is needed. This has been hard to achieve due to the proprietary nature of the measurement efforts, as is the case of [11, 12, 29], making it difficult for regulators to view measurement results from a harmonized and macroscopic scale. In this sense, several open measurement methodologies [30, 31] have been proposed with the goal of supporting the creation of inter-operable large-scale testbeds and advance a common approach on network performance characterization. The Internet Engineering Task Force (IETF) Large-Scale Measurement of Broadband Performance (LMAP) is currently working towards standardizing an overall framework for large-scale measurement platforms.

The MONROE platform complements the existing experimental platforms by providing unique features in the field of network-controlled mobile measurements. Three key aspects of MONROE that makes the platform unique are: repeatability and controllability of measurements for precise and scientifically verifiable results (even for the mobile scenarios), support for demanding applications such as web and video services and support for protocol and service innovation. These aspects sets up MONROE in an excellent position to advance the state-of-the-art measurement tools and platforms.

5.3 MONROE Approach and Key Features

MONROE's goal is to build a dedicated infrastructure for measuring and experimenting in MBB and WiFi (IEEE 802.11) networks, comprising both fixed and mobile hardware measurement nodes. The platform integrates 450 nodes scattered over four European countries (Italy, Norway, Sweden and Spain) and a backend system that collects the measurement results, offering tools for real-time traffic flow analysis as well as powerful visualization tools. We designed the MONROE nodes to be flexible and powerful enough to run most measurement and experiment tasks, including demanding applications like adaptive video streaming. The current MONROE node is an Accelerated Processing Unit (APU) with AMD 1 GHz dual core 64 bit processor and 4 GB DRAM. Each MONROE node connects simultaneously to three MBB networks through three MiFis using commercial grade mobile subscriptions. The nodes also provide WiFi connectivity[2] through a built-in dual band AC WiFi card. MONROE nodes have built-in support for collecting metadata such as cell ID, signal strength and connection mode. The nodes are equipped with GPS for tracking their location.

[2]The access points for WiFi will be provided when applicable for stationary nodes.

The MONROE platform allows external users to test their novel applications and services that run over MBB networks with WiFi connectivity. Through a user-friendly web client, external experimenters can schedule and deploy their own experiments on the MONROE nodes. Experimenters can use the MONROE platform to run measurements of different MBB providers at regular intervals over long time periods and under similar conditions.

The MONROE platform complements the existing experimental platforms such as RIPE Atlas [7] by providing unique features in the field of network-controlled mobile measurements. MONROE builds on the existing NorNet Edge (NNE)[3] [32] and extends its functionality, scale and coverage. The main features of MONROE are:

1) **Large-scale and wide geographical coverage**: MONROE is composed of 450 nodes that are widely distributed across Norway, Sweden, Italy and Spain, as we illustrate in Figure 5.1. MONROE is able to collect measurements under diverse conditions, from major cities to remote islands (including one node in Svalbard, in the Arctic). There is a dense deployment of nodes in a few main cities (e.g. Oslo, Stockholm, Madrid, Torino, etc.), giving a more detailed view of network conditions in urban areas.
2) **Mobility**: 150 MONROE nodes are deployed on trains and buses in order to cover both rural and urban areas. These nodes are instrumental to provide insights on the mobility characteristics of MBB.
3) **Multihomed**: Each MONROE node is connected simultaneously to three mobile broadband networks, which makes it possible to conduct a wide range of measurements and experiments that compare the performance of each network, or explore novel ways of combining resources from each network. Along with MBB networks, MONROE also provides WiFi connectivity to allow experimenting on different access technologies and explore methods such as traffic offloading.
4) **Flexible and powerful MONROE nodes**: The MONROE nodes are designed such that they are flexible and powerful enough to run most measurement and experiment tasks, including demanding applications like adaptive video streaming. Furthermore, MONROE enables experimenting novel services and applications on MBB networks by allowing configuration changes such as kernel modifications.

[3]NNE is currently in an operational state, with a functioning system for node management, deployment of experiments, handling of data etc. as well as real-time visualization of measurements (demo available at http://demo.robustenett.no).

Figure 5.1 Geographical distribution of MONROE Nodes. MONROE builds on the existing NorNet Edge (NNE) infrastructure, consisting of 200 dedicated operational nodes spread across Norway.

5) **Rich context information**: In addition to information about network, time and location for experiments, MONROE nodes have built-in support for collecting metadata from the externally connected modems, including cell ID, signal strength and connection mode.
6) **Open access**: MONROE is open to external users and makes it easy to access the system and deploy experiments on all or a selected subset of the nodes.
7) **Visualization and Open Data**: The MONROE platform has a measurement system that collects basic experiment results and then stores them in a database. Interested parties can then consume the measurement results through a real time visualization system. Furthermore, the results are provided as *Open Data* in regular intervals.

5.4 MONROE System Design

We designed the MONROE platform to make it easy for external experimenters to run their customized measurements. In this section, we expand on the MONROE system design and review the main building blocks and their functions. We illustrate the MONROE framework in Figure 5.2. Notably, MONROE not only allows to monitor and analyze the behavior of MBB network connections in real-time, but also to store measurement data jointly with metadata in the form of open data for offline analysis. The MONROE system comprises:

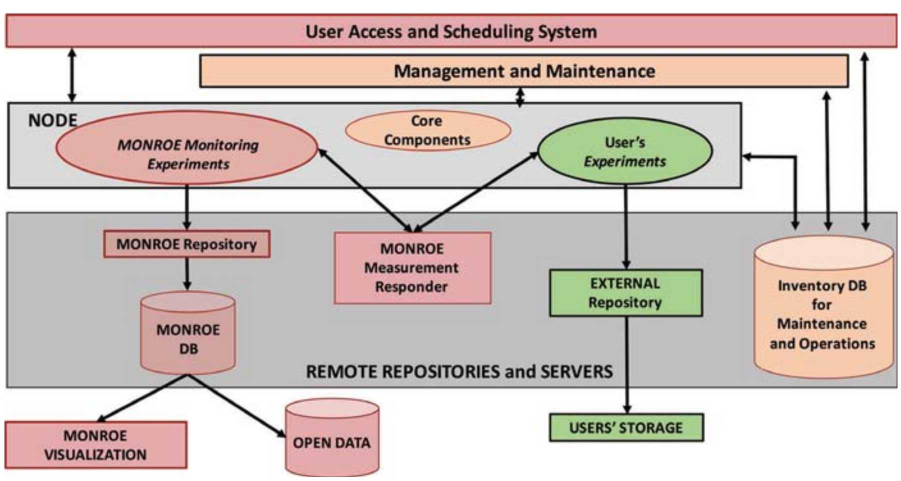

Figure 5.2 Building blocks of the MONROE system.

1. *User access and scheduling system*: The scheduling system handles the MONROE measurements through a user-friendly interface consisting of an AngularJS-based web portal. As part of the MONROE federation with the Fed4FIRE initiative of the European Commission[4], the user access follows the Fed4FIRE specifications in terms of authentication and provisioning of resources. The portal allows to access the MONROE scheduler, which is in charge of setting up the experiments without requiring the users to directly interact with the nodes (i.e., no login access to the node environment).
2. *Management and maintenance system*: The operations team uses this system to manage and maintain the MONROE testbed. It involves an *Inventory* that keeps all the information (e.g., the status of each node, status of different connections, location of the nodes, etc.) required for operations and maintenance. It also involves a *Monitoring Agent* that monitors and reports the health of the system (e.g., logging, performance monitoring, self checks for services etc.).
3. *Node modules*: The software on the measurement nodes includes the core management components and the set of experiments. The core components consist of the main software (watchdog, routing, network monitor, etc.) running on the node and make sure that the node is operational. An important core component is the Metadata Multicast, which is responsible for collecting and multicasting the metadata such as node status, connection technology and GPS. We provide a messaging API in order to relay real-time metadata to experimenters through ZeroMQ in JSON format.
   The experiments run in Docker[5] containers, which are running on a Debian Linux operating system. Containers can be described as lightweight virtualized environments and are particularly convenient since they allow agile reconfiguration and control of different software components. When external experimenters require kernel modifications to deploy their measurements, MONROE offers the possibility of using virtual machines within the node ecosystem. Experimenters can implemented and configure their measurements using any programming/scripting language, as long as the resulting experiment runs within these constraints.

[4]http://www.fed4fire.eu/
[5]http://www.docker.com

In order to monitor and assess the performance of MBB networks, MONROE continuously runs a basic set of experiments (*MONROE Monitoring Experiments* in Figure 5.2). Current deployed basic experiments include: continuous background measurements (e.g., ping to predefined servers), periodic bandwidth-intensive measurements, and a traffic analyzer developed in the mPlane project (`Tstat`). In Section 5.6.1, we expand on these measurements and analyze preliminary results. Apart from this, MONROE enables many other experiments for its external users (*User's Experiments* in Figure 5.2), which we further exemplify in Section 5.7.

4. *Repositories and Database*: The MONROE system supports external repositories to collect experimental data. Data transfer from nodes to the repositories is based on a set of agents that follow a publisher/subscriber model. We collect the results of the *MONROE Monitoring Experiments* in the MONROE repository and we subsequently import them to a centralized database for offline analysis. The database is based on a non-relational technology, oriented to time series analysis, and highly scalable to manage large volumes of data. We designed the database schema around the concept of experiments instead of physical nodes, with a clear distinction between experimental measurements and metadata. Several measurement responders we host in the MONROE backend act as measurement servers for certain experiments.
5. *Visualization and Open Data*: A near real-time visualization and monitoring tool enables stakeholders to access a graphical representation of the MONROE platform status in terms of deployment of the nodes, status of each device, as well as results of *MONROE Monitoring Experiments*. The results of selected measurements are provided as *Open Data* in regular[6].

5.5 Experiment Deployment

MONROE is an open platform for external users to experiment with MBB networks through active measurements. In this section, we detail the process an external user needs to follow in order to access the MONROE platform and we detail the MONROE components each experimenter interacts with. The work flow involves three main phases, as illustrated in Figure 5.3: Experiment Design, Testing and Experimentation.

[6]https://zenodo.org/collection/user-h2020_monroe

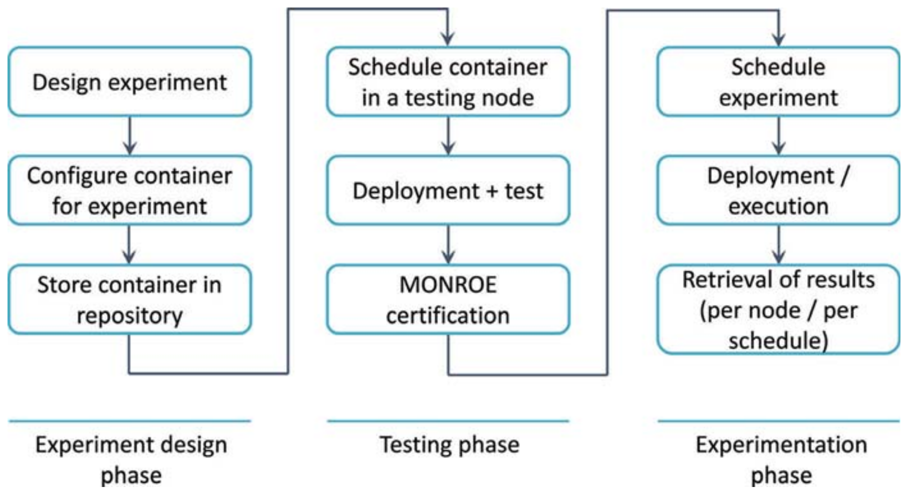

Figure 5.3 Experiment creation and deployment phases.

Experimenters have to define the measurements they want to obtain and decide how to implement them. Experiments run inside Docker containers, so they can consist of virtually any piece of software. During the testing phase, a MONROE administrator checks that the behavior of the container adheres to a set of minimum safety and stability rules; approved images are cryptographically signed and moved to our repository. Finally, the experimenter uses a web-based interface to schedule the experiment, selecting the number and types of nodes and suitable time-slots. Once the experiment is deployed and run, the results of experiments are automatically collected and transferred to a repository maintained by MONROE. Alternatively, experimenters can also choose to transfer/stream the results to their preferred location using their own independent solution.

Experiments can collect active and passive traffic measurements from multiple MBB networks. For active measurements the platform provides both standard/well-known tools (e.g., ping, paris-traceroute) and project-crafted ones. For passive measurements, it embeds tools such as `Tstat` [33] to analyze the traffic generated. Moreover, each node passively generates a metadata stream with modem and connectivity status, and the measurements of several embedded HW sensors (GPS, CPU usage, temperature, etc.). Experimenters can either subscribe their experiments to the stream in real-time or consult the database afterwards. Considering that experimenters can deploy any additional measurement tools, the set of possible measurements is flexible and open.

We provide User Access to the experimental platform via a web-based MONROE Experimenters Portal that enables users to schedule and run new experiments. The portal allows to access the MONROE scheduler, which is in charge of setting up the experiments without requiring the users to access the nodes. Since we federated MONROE within Fed4FIRE in order to build a large-scale, distributed and heterogeneous platform – authentication and provisioning of resources follows the Fed4FIRE specifications. In the following sections, we provide details on MONROE's federation with FED4FIRE, user authentication, experimenters portal and scheduler.

5.5.1 MONROE as a Fed4FIRE Federated Project

The Fed4FIRE Portal is a common and well-known tool where registered users can select and access an available testbed (e.g., the MONROE platform). The Fed4FIRE Portal is powered by MySlice software[7] and offers a directory of all FIRE testbeds, tools and links to project websites. In other words, the portal acts as an experimentation bridge to resources and their corresponding control tools.

To be able to join MONROE and run their experiments, the external users must first become familiar with the terminology and the tools of the Fed4FIRE federation and, in particular, with the MONROE project documentation. The available documentation of Fed4FIRE describes the federation of testbeds as a generic environment.

The user must apply for a Fed4FIRE account and download the corresponding required certificates, which should be associated with an existing MONROE experimentation project. The Fed4FIRE introductory documentation explains how to go through these particular steps. We note that the user must specify an already existing MONROE project, or alternatively, create a new one. In Section 5.5.2 we expand on how to complete the user authentication phase.

Once granted access to the platform, the user is recommended to follow and execute a MONROE tutorial, which describes those elements that are specific to the MONROE testbed, including the AngularJS client developed in the project for user access and experiment scheduling. Those users that plan to run measurement experiments in MONROE testbed should be familiar with the contents of the MONROE tutorial. To reserve the

[7]MySlice: http://myslice.info

resources for a specific experiment, the experimenter has to use the MONROE scheduler (Section 5.5.4), which can be accessed through the MONROE User Access client (Section 5.5.3). With the above, the experimenter can reserve the resources up to the limit granted to him/her by the MONROE consortium.

5.5.2 User Authentication

In this section, we describe the Fed4FIRE AAA policies and procedures, and how we adapt them to the MONROE project.

A federation is a collection of testbeds (or *"islands"*) that share and trust the same certification authorities and user certificates. Fed4FIRE realizes a federation of a large number of wired, wireless and OpenFlow-based testbeds principally located in Europe. Each island manages its resources using dedicated tools and can decide which kind of certificates (and from which authorities) it wants to accept. In this context, Fed4FIRE works with X.509 certificates to authenticate and authorize experimenters (users) on its testbeds. The authority which provides valid certificates in the Fed4FIRE federation is located at the iMinds infrastructure. The certification authority has the concept of Projects which bundle multiple users. Any user can requests for the creation of a new project, but it must be authorized by the Fed4FIRE administrators. Subsequently, the project responsible can approve new experimenters for that particular project, without prior approval from Fed4FIRE administrators.

MONROE shares and trusts the certificates generated by the iMinds authority, and therefore, is a member of the Fed4FIRE federation. We note that all the project functions and operations in MONROE depend on the user certificates, including resource reservation, measurements deployment and downloading experiment data. MONROE does not support other certification authorities or other federations (e.g., GENI).

Each partner in the MONROE consortium manages its own private project inside Fed4FIRE. Similarly, external institutions could have their own private projects upon request and approval by the MONROE Project Board. Individual researchers cannot join the MONROE testbed, as all the users must belong to at least one project (which corresponds to an institution that is managing it). However, each institution can easily invite new users and grant access to their respective projects offering the available resources which the MONROE administrators manage.

5.5.3 The Experimenters Portal (MONROE User Access Client)

Through the Experimenters Portal, verified external users can obtain access to the MONROE platform and deploy their measurements. After providing the necessary credentials to authenticate with the MONROE User Access client, the user can visualize a historic of all its experiments and check their current status (Figure 5.5). Clicking on any row of the table shows the details of the experiment selected.

Before scheduling new experiments, users can verify the current state of the MONROE resources. The *"Resources"* tab (Figure 5.4) allows experimenter to query all the existing resources in the MONROE platform and their time availability, using multiple filters if required.

In the *"New experiment"* tab, the user can create a new experiment and input the required parameters. The basic experiment details include the identifying name and the docker script to run the experiment. In the Experiment Size group, the user specifies the number of nodes required to run the experiment, and the desired characteristics of those nodes using filters that allows to select, e.g., the location of the nodes to use in the experiments, their hardware/software version, static or mobile nodes, testing nodes for preliminary/debugging tests, etc. Furthermore, the user can select the operator of interest and then define the maximum amount of data to be transferred per experiment over that interface/operator. This data limit is enforced during the experiments in order to avoid exceeding the mobile data quotas. In the Experiment Duration the user specifies the duration of the experiment by providing a starting and stopping date-time, or by clicking the "as-soon-as-possible" check box.

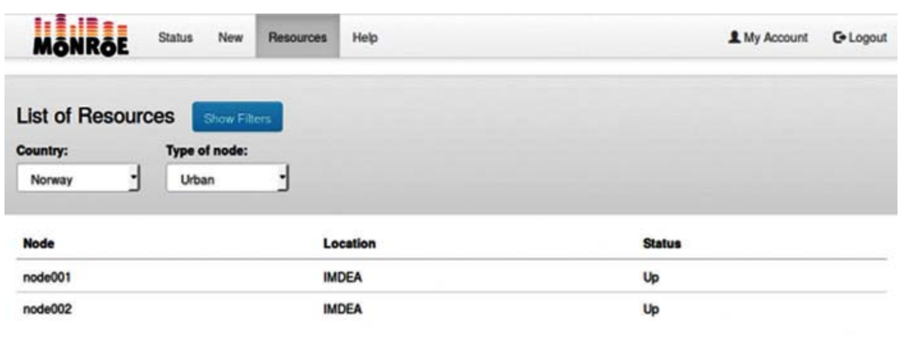

Figure 5.4 Resources availability in MONROE.

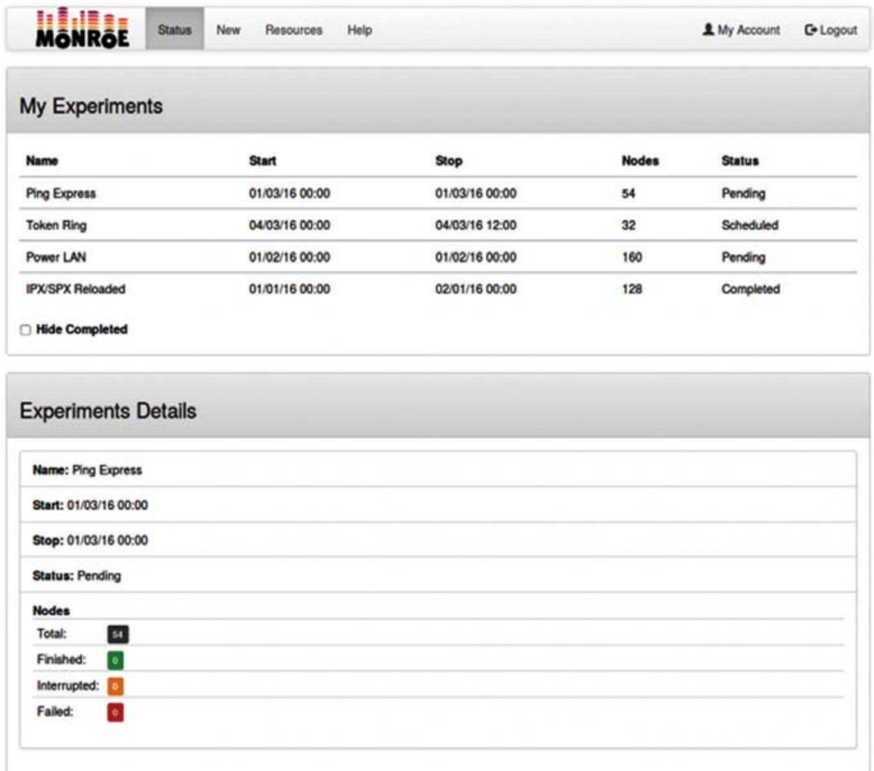

Figure 5.5 MONROE experiment status.

5.5.4 MONROE Scheduler

Through the MONROE User Access Client, the experimenters interacts with the MONROE Scheduler. The scheduler ensures that there are no conflicts between users when running their experiments and assigns a time-slot and node resources to each user.

In Figure 5.6 we present a schematic overview of the MONROE Scheduler functionality. We implement the MONROE Scheduler as a low-connectivity scheduling system which relies on the assumption that nodes are available, independent of short-time loss of connectivity. Due to the multihoming setup of the MONROE nodes, they may contact the scheduler from different addresses, possibly with provider-dependent modifications and filters. The Scheduler consists of two components – the *scheduling server* running in a central, well-known location and the *scheduling client* running on the nodes (Figure 5.6).

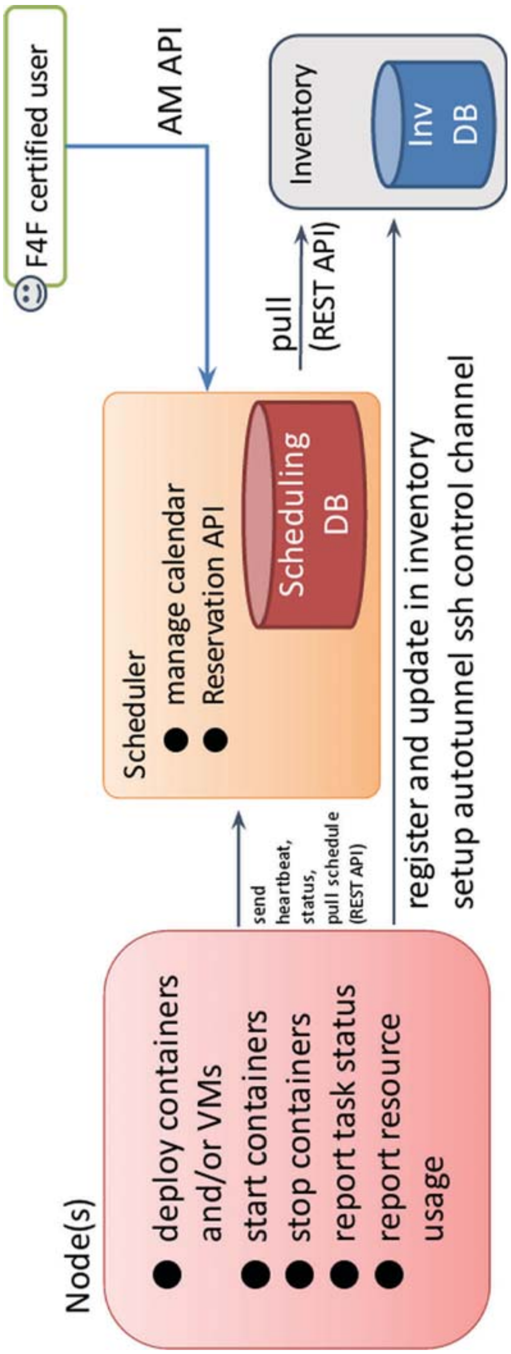

Figure 5.6 Scheduling system.

The scheduling server:

- takes care of the experiment schedule and resolves conflicts
- assigns roles to authenticated users
- provides a REST API to users and nodes to query and edit scheduling status
- provides an XML-RPC API compatible with the Fed4FIRE AM API definition

The scheduling client:

- sends a regular heartbeat and status to the scheduling server
- fetches the experiment schedule for the current node
- downloads, deploys, starts and stops scheduled experiments

Authentication to the server is based on X.509 client certificates. Users, administrators and nodes all authenticate using this mechanism and use the same scheduling API. By importing the Fed4FIRE certification authority certificate, users may authenticate using their Fed4FIRE credentials.

Due to the connectivity constraints especially of mobile nodes, deployment of experiments on the node is not immediate. Download and deployment of experiments will take place as early as possible within the constraints of available space on the node. The node will report a successful deployment to the scheduler and schedule the start and stop times for the experiment container internally. Changes in the schedule are propagated to the node whenever possible.

The MONROE Scheduler implements the procedures and policies we have defined to guide the MONROE experimentation. These include, but are not limited to:

- The scheduler allows booking of fixed time slots for each measurement experiment.
- Priority is defined by the first-come first-serve principle, while the consortium will monitor fairness.
- If an experiment is marked as *exclusive*, only one experiment may run at a given time on a node.
- If an experiment is marked as *active*, one such experiment may run at a given time on a node, while allowing passive experiments.
- If an experiment is marked as *passive*, a given number of such experiments may run at a time. No traffic may be generated by the experiment.
- User experiments may be scheduled as periodic, continuous, or one-time.

- Only experiments for which a time slot has been booked in advance may be run.
- Nodes may be of different types (static, mobile, urban, rural, certain country, etc. . .) defined by the MONROE project. Booking requests can select to use or reject these filters.
- A booking over several nodes or several time periods is treated as atomic (i.e., if one of the booking periods or nodes is unavailable, the entire booking is rejected). Several bookings over different nodes or time periods may be linked to an atomic unit.

In order to determine the resource requirements, each user needs to schedule its experiment to first run on the testing nodes (Testing Phase in Figure 5.3). This step allows us to monitor the resource usage of each experiment. If the usage is within defined constraints, the MONROE administrators move on to approve the user experiments by means of a cryptographic signature. Only then, the experiment image is cleared to be scheduled on regular nodes.

The scheduling process on the node (Deployment Phase in Figure 5.3) defines three actions: (i) deployment, (ii) start and (iii) stop of the experiment. The deployment step may take place at any time before the scheduled start time, and should finish before the experiment starts. In this step, the scheduler reserves the requested resources and loads the experiment image onto the nodes. During the start process, the scheduler sets the resource quotas and starts using the experiment image a container system where experiments will run. The stop action notifies the experiment of its impeding shutdown, then removes the container after a short grace period. Measurement results may be stored on disk, and will be transferred during and after the termination of the experiment as connectivity allows.

5.6 Network Measurements and Analytics with MONROE

The MONROE platform continuously runs a set of basic measurements with the purpose of characterizing the state of the MBB providers in Europe. Interested parties can consume the data through the MONROE visualization GUI, thus making MONROE a solution for near real-time network performance monitoring. In Figure 5.7, we show a snapshot from the MONROE monitoring interface tracking a node in terms of both RTT and signal strength. Alternatively, we provide the measurement results as open data which external users can access and use for running network analytics.

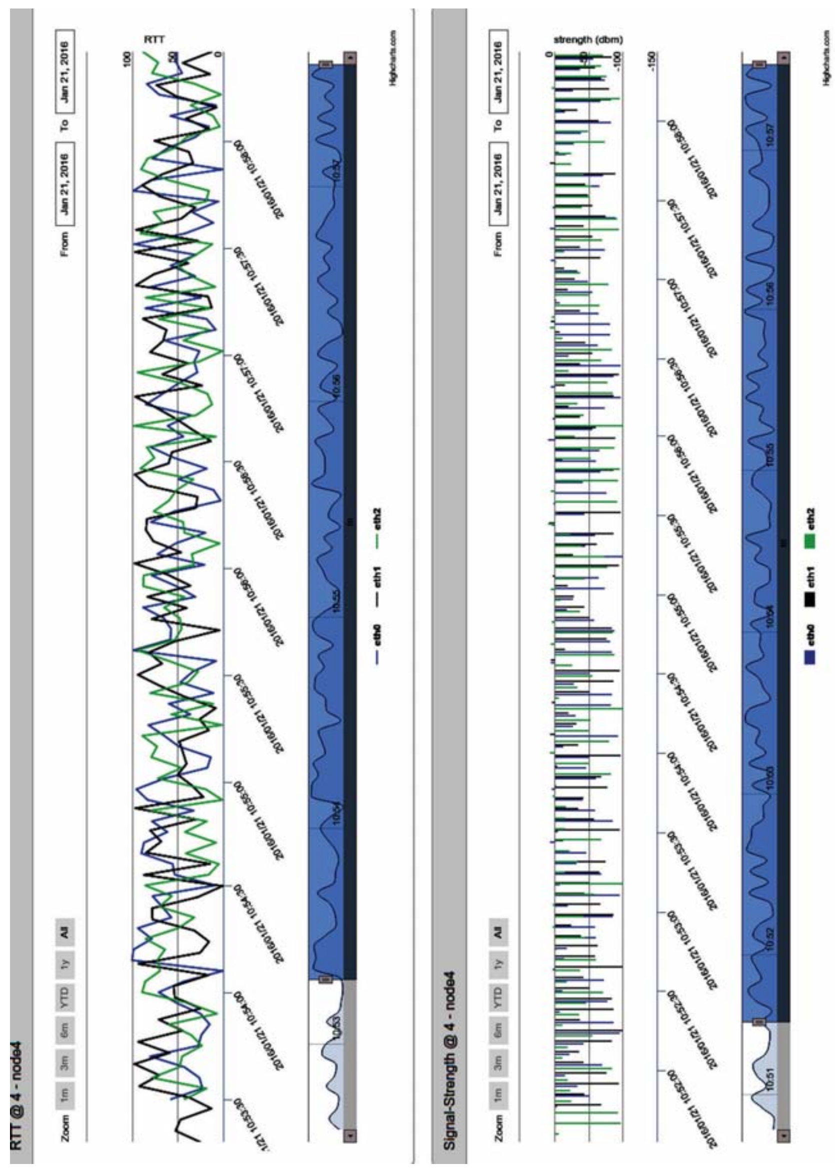

Figure 5.7 MONROE visualization GUI snapshot for RTT and signal strength monitoring in near real time.

5.6.1 MONROE Monitoring Experiments

The MONROE Monitoring Experiments currently include (but are not limited to) i) continuous ping measurements towards a fixed target in Sweden, ii) a simple bulk data download, and iii) web browsing performance measurements. The MONROE nodes also continuously run Tstat [33], a passive monitoring tool developed within the mPlane project [34]. Tstat extracts information from the flow of packets being transmitted and received by each node. This facilitates the use of the MONROE platform as an analytic tool for troubleshooting and root-cause analysis. In this section, we report preliminary measurement results illustrating the capabilities of the platform towards performance monitoring and network analytics.

a) RTT Measurements: Each MONROE node runs a ping measurement every second on each active interface against the same target measurement server we host in the MONROE backend in Sweden. Figure 5.8 shows the violin plot for the RTT samples we collected during one week (from the 8th of July until the 15th of July 2016) from 30 stationary nodes connected in total to 7 different operators in 3 countries. Each “violin” shows the probability density of the RTT at different values, the higher the area, the higher

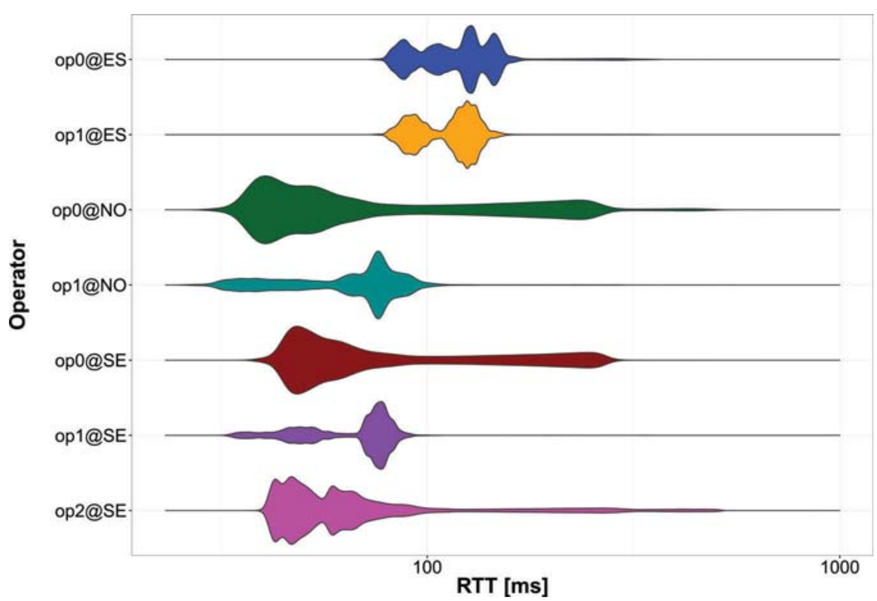

Figure 5.8 Violin plots of the RTT measurements for different operators in Spain (ES), Norway (NO) and Sweden (SE).

the probability of observing a measurement in that range. We observe that the RTT measurements exhibit typically a multimodal distribution, corresponding to different access delays faced by different radio access technologies (e.g., 3G/4G).

The results are intuitively expected: nodes in Norway and Sweden that are closer to the target measurement server (which we host in the MONROE backend in Sweden) exhibit lower delay than the nodes in Spain. However, the variance of the measurements is much higher than in fixed networks, showing that MBB introduces complexity even for basic tests, such as RTT monitoring. Given that the ping experiment is running continuously, some of this variation can be due to interactions with other experiments running on the MONROE nodes. The repetitive measurements allow us to track this key parameter in time and capture the experience of customers using mobile subscriptions similar to those active on the MONROE node. By analyzing the RTT time series, we plan to further identify delay trends and correlate them with the time of the day, the geolocation of the measurement node and the rich context information we collect from the devices (e.g., RAT changes and variations in the signal strength). This uniquely enables us to work towards understanding congestion patterns in the networks.

b) Download Capacity Measurements: In Figure 5.9, we illustrate downlink throughput measurement results. Every two hours, we schedule the download of a 50 MB file on 30 stationary MONROE nodes on all interfaces corresponding to seven different MBB operators from an HTTP server we host in the MONROE backend in Sweden. Running in the background, Tstat analyzes this traffic and generates different key performance metrics, including download throughput and the RTT from the client to the server. Plots in the top row of Figure 5.9 show the CDF of the download throughput, while plots on the bottom show the evolution over three days of experiments (from the 22nd of July until the 24th of July) of the average RTT as observed by Tstat during the transfer. We note that performance varies wildly among countries, among operators within the same country and over time.

As expected, nodes in Spain located further away from the measurement server display a higher RTT than the nodes in Norway or Sweden. Also, we see a clear separation between the RTT we measure in Norway for the two operators. Based on further analysis we perform with Tstat, we identify the presence of a non-transparent proxy in the network of operator *op1*. We further note the impact of the web proxy when monitoring the goodput metric for both operators in Norway: *op1* benefits from the proxy and displays a higher goodput than *op0*.

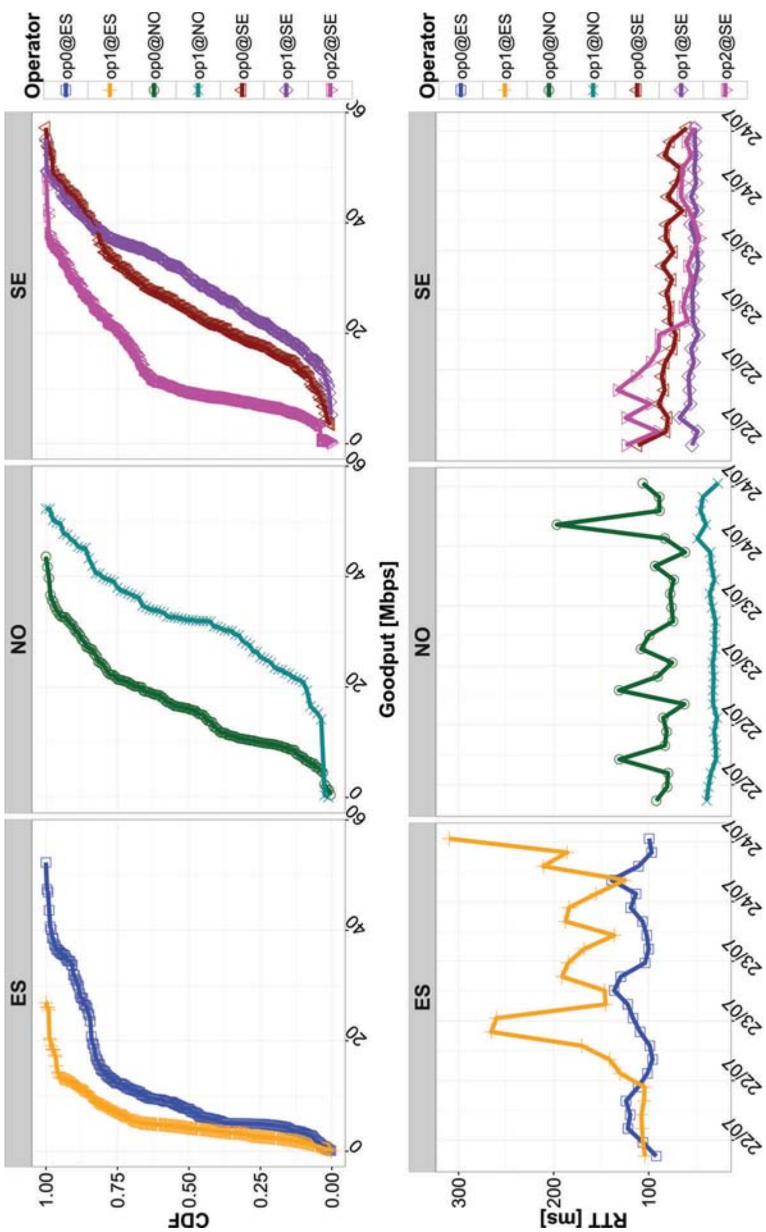

Figure 5.9 CDF of the download goodput (*top*) and average RTT (*bottom*) for different operators in different countries.

c) Web Browsing Performance: Aside from the basic measurements that run continuously on the measurement nodes, we design and periodically schedule a specific experiment to gauge web browsing performance across multiple MBB providers in different countries. Each MONROE node connects on each interface to two different websites[8], which we chose based on their popularity in the Alexa ranking, but also based on their different appearance and rendering style. As part of the experiment design, the web performance test breaks down the times used for different phases in a web transaction at each interface of the MONROE node: time to resolve the DNS name, time to connect to web server and time to download the web content and all its objects (including elements generated by javascript). Also, the web performance test tracks several other metrics to describe the web browsing activity and the target website, including number of DNS iterations, number of HTTP redirects, number of HTTP elements or HTTP download size.

In Figure 5.10, we illustrate the CDF of the complete page load time and the CDF of the average time to first byte of content broken down per country and per website we target. We observe significant variance in both metrics. This happens because some pages (e.g., *en.wikipedia.org*) consist of fewer objects, and therefore can complete faster. The median object counts per web page are 69 for *www.bbc.com* and 14 for *en.wikipedia.org*. Other pages take longer to download because they have several objects that may be fetched from multiple servers. Also, for the Spanish operators, we detected multiple number of DNS iterations for *www.bbc.com*, thus partially explaining the higher TTFB metric compared to other operators in Norway and Sweden.

Discussion: While these experiments are preliminary, they clearly show the need of experimental investigation to understand 3G/4G network and application performance. The MONROE platform offers researchers the unique opportunity to run and repeat experiments to provide evidence of complicated phenomena.

5.6.2 Network Analytics with MONROE

One of the main targets of the MONROE platform is to provide experimenters a rich dataset of key mobile broadband metrics, from which different stakeholders can further extract the information of interest regarding the performance and reliability of MBB networks. To measure the network in a reliable and fair way, it is crucial to identify the metrics that accurately capture the performance

[8]The two websites we target are *"www.bbc.com"* and *"en.wikipedia.org"*.

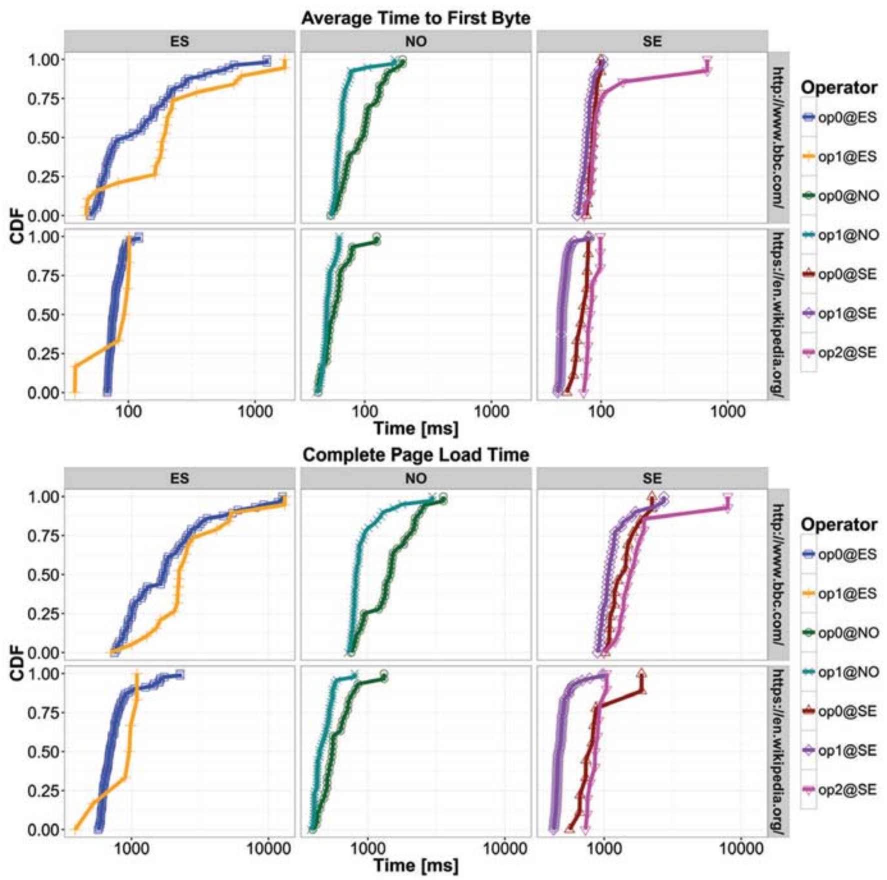

Figure 5.10 Web performance results: the Average Time to First Byte and the Complete Page Load Time for operators in Spain (ES), Norway (NO) and Sweden (SE) for two target websites www.bbc.com and en.wikipedia.org.

and the conditions under which we evaluate these metrics. Different stakeholders have different requirements on the metrics supported by the MONROE platform. For example, on the one hand, regulators need connectivity, coverage and speed information collected from a third-party, independent platform to monitor whether operators meet their advertised services, and as a baseline for designing regulatory policies. On the other hand, operators are interested in time series reporting of operational connectivity data to identify instability and anomalies. Furthermore, application developers need to cross-check QoS parameters against the behavior of the underlying network to design robust services and protocols. From the above considerations, it is clear that the collection of data cannot be limited to transmission and packet-level statistics, but there is an obvious need for rich metadata to be associated with the performance and reliability measurements.

The network metadata enables MONROE to capture the network context under which we measure the key performance metrics. The parameters we report include but not limited to provider name, radio access technology (RAT) type, RAT-specific parameters (e.g., RSRP, RSRQ, RSSI) and network connectivity status. Network metadata is crucial not only for coverage information but also during the analysis of the measurements in order to understand the underlying factors that affect the performance.

a) Mimicking Drive Tests for Mobile Coverage: One essential aspect when monitoring MBB providers is characterizing the coverage offered to unveil complex patterns of different radio access technologies (RATs) in an area. Network operators regularly test different network parameters of their deployed infrastructure for network benchmarking, optimization, troubleshooting and service quality monitoring. This is usually done via drive-testing where measurements are either collected by a vehicle with an embedded GPS device and other measurement equipments e.g. a laptop or by using mobile phone with an engineer roaming around the streets and roads of a region so that to have an end-user experience. However, there are major drawback to this approach, mainly the high cost it entails in terms of time and labor, and also that it does not cover most of the region where there are customers. The mobile MONROE nodes (placed on public transport vehicles) enable mimicking the drive tests measurements resulting in a dataset similar to the ones operators work with. Piggy-backing network measurements onto public transportation vehicles via MONROE offers additional benefits, including ensuring repeatability of drive runs on the same route, in similar busy-hour conditions, since the MONROE node is active in the times when the trains or buses carry passengers to their destinations. This approach emerges as a cost-effective alternative to the drive test performed by operators, with the added perk of allowing other parties, including public transport companies, to assess and compare the MBB coverage along their infrastructure at a zero added cost.

In Figure 5.11, we illustrate the measurement location from the mobile nodes active aboard trains inside Oslo are in Norway. We color-code the data points to show the radio access technology we read from the modem connected to one of the operators we measure. We observe that majority of time the node has 3G coverage and intermittent 4G coverage.

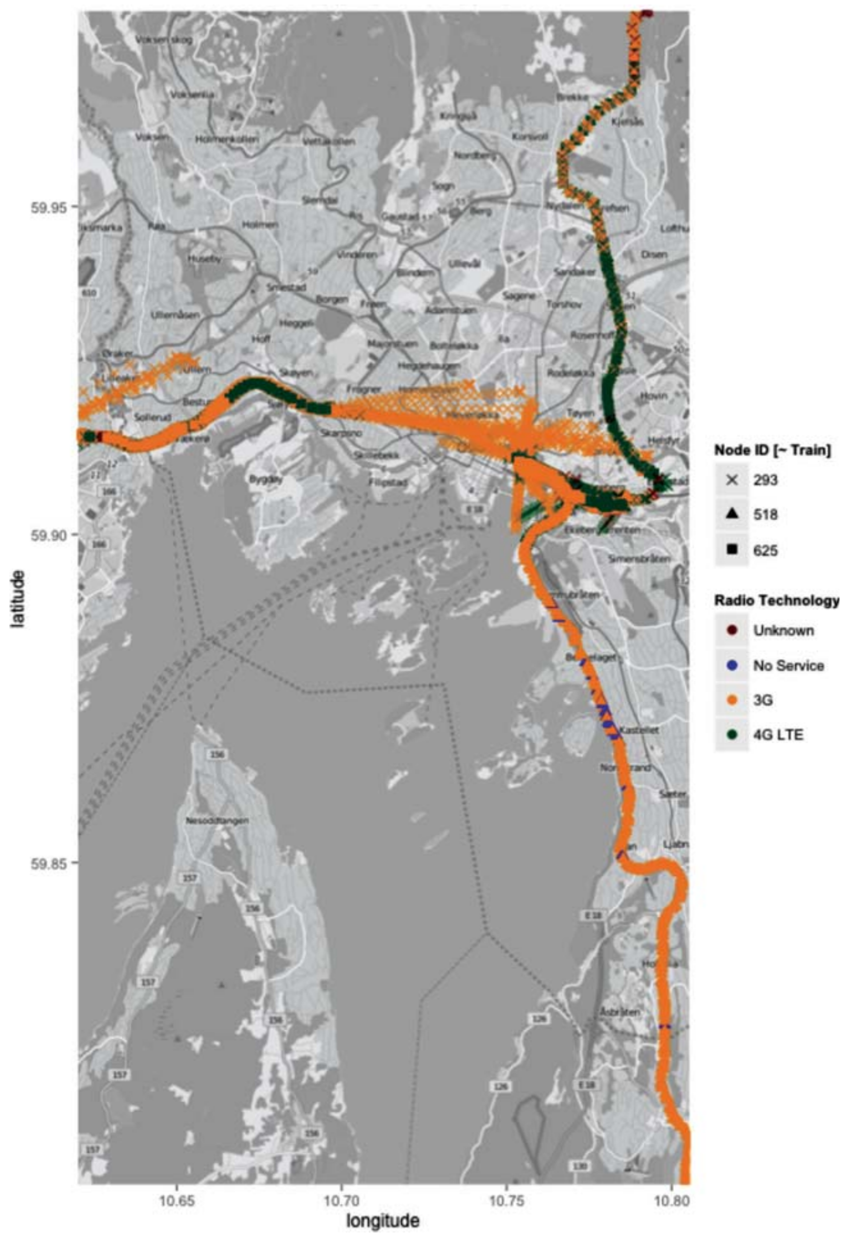

Figure 5.11 Coverage reading from MONROE nodes operating aboard trains in Oslo, NO.

5.7 User Experiments

Along with being a near real-time monitoring and benchmarking platform, MONROE is an open platform for experimentation with MBB networks. Below, we list a set of representative examples that MONROE users are currently curating. This serves to further illustrate the value of the MONROE platform and the variety of experiments it can accommodate.

a) Service Oriented Quality of Experience: A first dimension to explore comes from the great interest in how users perceive individual services and applications over different terminals (e.g., mobile phones, tablets, and computers). The recent proliferation of user-centric measurement tools such as Netalyzr [6] to complement available network centric measurements validate the increasing interest in integrating the end user layer in network performance optimization. MONROE enables experimentation with essential services and applications, including video streaming, web browsing, real-time voice and video, and file transfer services. The service oriented measurements give a good bases for investigating the mapping from Quality of Service to Quality of Experience.

b) Protocol Assessment: A second dimension to explore consists in the assessment of existing and new protocols in MBBs on a scale that was previously not possible. The large availability of experimental resources in MONROE is well suited to assess networked applications under a wide range of network conditions, while still giving experimenters strong control of the testing environment. Furthermore, the multihoming aspect of MONROE nodes makes it ideal for experimenting with protocols that exploit multiple connections opportunistically, e.g., in parallel or by picking the one with the best available service to increase robustness and performance, or to achieve the best cost-performance ratio. Examples of such protocols and services include, but are not limited to, Multipath TCP, Device-to-Device for offloading or public safety applications, portable video streaming services or e-health services.

c) Middlebox Impact: Another significant use case for MONROE is related to the use of middleboxes. These can range from address and port translators (NATs) to security devices to performance enhancing TCP proxies. Middleboxes are known to introduce a series of issues and hinder the evolution of protocols such as TCP. Therefore, measuring and understanding their behavior is essential. Since middleboxes of different types are ubiquitous in MBB networks, a platform such as MONROE offers an excellent vantage point from which to observe and characterize middlebox operation in real world deployments.

d) Knowledge Discovery and Network Analytics: Beyond mere service and protocol assessment, MONROE offers the possibility to develop mechanisms to augment network performance by learning from measurements. This use case involves post processing of data, to deepen the understanding of network behaviors. The goal is to identify causalities and correlation of different parameters that can individually or collectively affect the performance and reliability of the network. In order to identify unexpected data patterns that deserve attention, one should go beyond data-mining and correlation approaches, and rather use knowledge description techniques, such as the Kolmogorov complexity method [35] or the minimum description length theory [36]. Such approaches are beneficial for different stakeholders including operators, vendors, developers and service providers. Therefore, we envision MONROE to have a significant impact on different sectors of industry through these knowledge discovery approaches, while helping to improve the performance of their products leading to a better user experience for the end users.

5.8 Conclusions

In this chapter, we introduce the MONROE platform: an open and industry-grade platform for MBB measurements and experiments. The MONROE platform enables accurate, realistic and meaningful assessment of the performance of MBB networks by continuously monitoring these networks via active testing (e.g., delay test, web performance test, download speed test) and context metadata collection (e.g., connection mode, signal strength parameters). Furthermore, MONROE provides the perfect setting to test novel services and protocols thanks to its flexible and powerful nodes with multihoming support. In this chapter, we showcase the monitoring capabilities of the platform by analyzing preliminary performance measurement results. We further describe various examples of experiments that are supported by the platform in order to illustrate the unique features of the MONROE platform.

We argue that mobile measurements over operational networks are essential to understand the fundamental characteristics of mobile ecosystem as well as to establish the quality of end user's experience for different services. Such information is valuable to many different stakeholders including operators, regulators, policy makers, consumers, society at large, businesses whose services depend on MBB networks, researchers and innovators. For example, MONROE measurement results provide insights that can enable operators with more accurate radio resource and infrastructure planning, more cost-efficient investments, and better network utilization. Operators can also explore differentiated and specialized services, as well as their requirements and impact on applications. Application developers for mobile devices can use the platform to test various applications and services over MBB. With better knowledge about MBB and the ability to test services, MONROE will contribute to service providers innovating more and realizing innovative services. Internet of Things and smart city services will lead in this direction as more vertical specific applications and services will be developed along with the evolution towards 5G. Due to multihomed support, innovations regarding network selection, handover and aggregation can be developed to make applications more robust with better adaptability and increased quality; for this, multipath TCP and Device-to-Device communications are instrumental. These are a few examples of the opportunities in the MBB field that requires extensive research efforts from both industry and academia, and the MONROE platform with its unique features is the key enabler to achieve them.

References

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper*, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, Cisco Systems, Inc., February 2016, accessed: 2016-08-01.

[2] networld2020, "Service level awareness and open multi-service internetworking – principles and potentials of an evolved internet ecosystem," 2016.

[3] FCC, "2013 Measuring Broadband America February Report," FCC's Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, Tech. Rep., 2013.

[4] Tektronix, "Reduce Drive Test Costs and Increase Effectiveness of 3G Network Optimization," Tektronix Communications, Tech. Rep., 2009.

[5] OOKLA, "http://www.speedtest.net/."

[6] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzr: illuminating the edge network," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, 2010, pp. 246–259.

[7] R. Atlas, "https://atlas.ripe.net/."

[8] S. Sundaresan, S. Burnett, N. Feamster, and W. De Donato, "Bismark: a testbed for deploying measurements and applications in broadband access networks," in *2014 USENIX Conference on USENIX Annual Technical Conference (USENIX ATC 14)*, 2014, pp. 383–394.

[9] PlanetLab, "https://www.planet-lab.org/"

[10] Z. Wang, Z. Qian, Q. Xu, Z. Mao, and M. Zhang, "An untold story of middleboxes in cellular networks," in *Proc. of SIGCOMM*, 2011.

[11] "OpenSignal^{TM}," http://opensignal.com

[12] "RootMetrics^{TM}," http://www.rootmetrics.com/

[13] "MobiPerf," http://www.mobiperf.com

[14] A. Nikravesh, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh, "Mobile Network Performance from User Devices: A Longitudinal, Multidimensional Analysis," in *Procs. of PAM*, 2014.

[15] M. Molinari, M.-R. Fida, M. K. Marina, and A. Pescape, "Spatial interpolation based cellular coverage prediction with crowdsourced measurements," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*. ACM, 2015, pp. 33–38.

[16] M. K. Marina, V. Radu, and K. Balampekos, "Impact of indoor-outdoor context on crowdsourcing based mobile coverage analysis," in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*. ACM, 2015, pp. 45–50.

[17] A. Le, J. Varmarken, S. Langhoff, A. Shuba, M. Gjoka, and A. Markopoulou, "Antmonitor: A system for monitoring from mobile devices," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*. ACM, 2015, pp. 15–20.

[18] Z. Wang, F. X. Lin, L. Zhong, and M. Chishtie, "How far can client-only solutions go for mobile browser speed?" in *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012, pp. 31–40.

[19] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "Yomoapp: A tool for analyzing qoe of youtube http adaptive streaming in mobile networks," in *Networks and Communications (EuCNC), 2015 European Conference on*. IEEE, 2015, pp. 239–243.

[20] A. Nikravesh, H. Yao, S. Xu, D. Choffnes, and Z. M. Mao, "Mobilyzer: An open platform for controllable mobile network measurements," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2015, pp. 389–404.

[21] J. Sommers and P. Barford, "Cell vs. WiFi: On the Performance of Metro Area Mobile Connections," in *Proc. of IMC*, 2012.

[22] W. L. Tan, F. Lam, and W. C. Lau, "An empirical study on 3G network capacity and performance," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. IEEE, 2007, pp. 1514–1522.

[23] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3G cellular data network," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1341–1349.

[24] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck, "An in-depth study of LTE: Effect of network protocol and application behavior on performance," in *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, ser. SIGCOMM '13. New York, NY, USA: ACM, 2013, pp. 363–374. [Online]. Available: http://doi.acm.org/10.1145/2486001.2486006

[25] P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks," *SIGMETRICS Perform. Eval. Rev.*, Vol. 41, No. 2, pp. 44–46, Aug. 2013. [Online]. Available: http://doi.acm.org/10.1145/2518025.2518033

[26] F. Fund, C. Wang, Y. Liu, T. Korakis, M. Zink, and S. Panwar, "Performance of dash and webrtc video services for mobile users," in *Packet Video Workshop (PV), 2013 20th International*, Dec 2013, pp. 1–8.

[27] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 2012, pp. 225–238.

[28] N. Thiagarajan, G. Aggarwal, A. Nicoara, D. Boneh, and J. P. Singh, "Who killed my battery?: analyzing mobile browser energy consumption," in *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012, pp. 41–50.

[29] *Methodology and technical information relating to the SamKnows™ testing platform – SQ301-002-EN*, SamKnows™, 2012.

[30] M. Bagnulo, P. Eardley, T. Burbridge, B. Trammell, and R. Winter, "Standardizing Large-scale Measurement Platforms," *SIGCOMM Comput. Commun. Rev.*, Vol. 43, 2013.

[31] S. Sen, J. Yoon, J. Hare, J. Ormont, and S. Banerjee, "Can they hear me now?: A case for a client-assisted approach to monitoring wide-area wireless networks," in *Proc. of IMC*, 2011.

[32] A. Kvalbein, D. Baltrūnas, J. Xiang, K. R. Evensen, A. Elmokashfi, and S. Ferlin-Oliveira, "The Nornet Edge platform for mobile broadband measurements," *Elsevier Computer Networks special issue on Future Internet Testbeds*, 2014.

[33] A. Finamore, M. Mellia, M. Meo, M. M. Munafo, P. D. Torino, and D. Rossi, "Experiences of internet traffic monitoring with tstat," *IEEE Network*, Vol. 25, No. 3, pp. 8–14, May 2011.

[34] P. Casas, P. Fiadino, S. Wassermann, S. Traverso, A. D'Alconzo, E. Tego, F. Matera, and M. Mellia, "Unveiling network and service performance degradation in the wild with mplane," *IEEE Communications Magazine*, Vol. 54, No. 3, pp. 71–79, March 2016.

[35] M. Li and P. M. Vitnyi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed. Springer Publishing Company, Incorporated, 2008.

[36] P. D. Grünwald, *The Minimum Description Length Principle (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.