Experimental Performance Evaluation of WebRTC Video Services over Mobile Networks

Mohamed Moulay  
IMDEA Networks Institute, Madrid, Spain  
and University Carlos III of Madrid, Spain  
mohamed.moulay@imdea.org

Vincenzo Mancuso  
IMDEA Networks Institute, Madrid, Spain  
vincenzo.mancuso@imdea.org

Abstract—With the rapid growing market for smartphones, and user’s confidence for immediate access to high-quality multimedia content, the delivery of video over wireless networks has become a big challenge. WebRTC is a new and evolving standard that has been developed specifically to meet this demand and enable a high-quality experience for mobile users of real-time communication services. However, little systematic experimental studies have been carried out so far to assess the service experienced by users in a realistic mobile setting. In this work, we describe measurements collected from a WebRTC implementation operated from real mobile nodes within the pan-European MONROE platform. Using data from application and network, in different wireless environments, we experimentally investigate and compare the performance of WebRTC for static and mobile cellular users of several networks across Europe.

Index Terms—WebRTC; MBB; Experiments.

I. INTRODUCTION

Web Real-Time Communication (WebRTC) proposes to easily integrate video services in web browsers, based on local tools [1]. In turn, such tools are based on well-known web technologies, able to simply integrate audio, video, and data transfer operations of the real-time communication protocol (RTC) into a normal webpage. The WebRTC project\(^1\) was first introduced by Google as an open source project, and then other software developers and telecom vendors joined, which has led to integration of WebRTC into commercial browsers like Chrome, Opera and Firefox [2], [3].

Since offering video services and multimedia channels is a killer application for Mobile broadband (MBB) networks, WebRTC and similar projects impose stringent quality requirements on such networks, that are nowadays evolving from 4G to 5G under the pressure of a steadily increasing number of mobile users [4]. Therefore, there is now a strong need for objective information about MBB performance and reliability to support video and multimedia mobile services. Thus, various initiatives have arisen, among which the US FCC’s Measuring Broadband America initiative [4] and MONROE [5], to monitor and assess MBB performance.

We focus on WebRTC performance figures in mobile environments, for which so far little experimental work exists. In fact, other works on assessing WebRTC performance figures are currently sprouting, but they have so far only investigated basic properties in controlled environments. For instance, the authors of [6] used a cloud-engineered automatic testing tool for WebRTC, although they have not tested the service offered by mobile operators and core networks. Similarly, the authors of [7] have experimentally tested one-to-many communications over WebRTC (namely “simulcast”), although their experiments are limited to a gigabit LAN environment.

In contrast, for our work, we leverage on a large-scale on-line measurement platform and focus on users connecting through MBB networks only. Specifically, we use the above mentioned MONROE platform, which has been designed and is currently operated in the frame of a European project aimed at providing multi-homed, independent, large-scale monitoring and evaluation of performance for mobile broadband networks in heterogeneous environments. Acquiring access to this platform allows for the deployment of vast measurement setups to collect data from operational MBB networks in various European countries. Differently from other approaches based on operator-driven quality-assessment campaigns [8], [9], or on traditional drive-by tests [10], MONROE offers an open platform for repeatable and traceable experiments. Besides, it offers open access to collected data, which refer to multiple operators, and includes device-level metadata, which is the key to use and possibly filter results without raising user’s privacy concerns. Therefore, this platform offers much richer data than what can be offered by crowdsourcing initiatives like, e.g., Netalyzer [11] and Haystack [12].

In this paper, we use the MONROE platform to investigate session-related performance statistics linked to the use of an off-the-shelf, WebRTC-based streaming application. This application enables streaming videos in real time with high quality using web browsers that support WebRTC (e.g., Chrome, Firefox) [1]. When using Google Chrome, data from both the sending and receiving parties in a WebRTC-based telemeeting, can be gathered via the WebRTC internals page,\(^2\) thus making it possible to get a complete overview of the stream in the mobile nodes. Such session-related statistics may help to identify root causes and track the origins of performance issues in video streaming, so to understand how these technical factors impact Quality of Service (QoS) offered by the network and Quality of Experience (QoE) enjoyed by

\(^{1}\)http://www.webrtc.org/

\(^{2}\)chrome://webrtc-internals/
The main API highlights are as follow:

- **GetUserMedia**: it provides with agile access to user media such as microphones, cameras and display.
- **RTCDataChannel**: it authorizes data transfer through peer-to-peer channels.
- **RTCPeerConnection**: it is responsible for setting up a direct connections between two WebRTC applications, which allows data channels and media streams to be carried.

This API represents the base of any WebRTC application. Besides, it is possible to add plugins to exploit WebRTC or to empower it with, e.g., encryption and security features [15].

### B. Protocols and Communication Services

Web browsers have to support several communication protocols and communication services to enable WebRTC. As shown in Fig. 1, this includes mechanisms such as signaling, session establishment, transport and security.

For transport, WebRTC commonly uses UDP with DTLS security, which is a TLS extension for unreliable datagram transport. However, WebRTC gives the option to use TCP with TLS as well. In addition, it uses the Stream Control Transport Protocol (SCTP) and the Secure Real-Time Transport (SRTP) to control media channels and handle encryption keys. SCTP is capable of multiplexing multiple application data streams and provides reliable delivery of UDP datagrams. Hence, a peer-to-peer secured media path can be established by leveraging SCTP and SRTP.

For what concerns the session setup and the negotiation of media features and connection parameters, WebRTC uses SIP or XMPP with parameters conveyed through the Session Description Protocol (SDP). In addition, WebRTC uses the JavaScript Session Establishment Protocol (JSEP) as session setup mechanism, which endorses secured signaling, and encrypted data transport over either DTLS/UDP or TLS/TCP. It also uses signaling servers that help initiate peer-to-peer multimedia capabilities handshaking and establish connections.

WebRTC also includes mechanisms for firewall and NAT traversal. Specifically, it inherits from the VoIP world the Interactive Connectivity Establishment (ICE) framework. ICE helps to use TURN/STUN, which in turn eases WebRTC peers to test and collect preferred “candidates of communication” options, i.e., it sorts the known transport addresses of a media destination to which is possible to attempt connection. Indeed, carefully selected communication candidates are the key to maximize the chance of success.

The most recent specifications on WebRTC have been released in November 2017 with the W3C EditorDraft “WebRTC 1.0: Real-time Communication Between Browsers.”

### III. THE MONROE MEASUREMENT PLATFORM

Since we use MONROE for our WebRTC measurements, here we present a short overview on the measurement platform. The main idea behind MONROE is to deploy hundreds...
Fig. 2: A MONROE node pair with three LTE Cat6 modems and one WiFi adapter.

Fig. 3: MONROE platform components and information flow.

---

**maintenance and Management:**
This block is a subsystem used by the maintenance team to administer and maintain MONROE nodes with ease. It relies on a DB with a friendly interface to check node connectivity status and stats.

**Node:**
Measurement nodes are dual-apu2 machines with a Linux Debian based operating system. They include core components (e.g., for handling routing and connectivity, for monitoring the software behavior, etc.) plus a set of Linux Docker containers in which experiments are trigged in isolation. Moreover, some containers run constantly in the background to collect results from “default” experiments, e.g., to collect statistics on bandwidth throughput, delay, and gather metadata such as node GPS coordinates. The main idea behind using containers is that it allows for handy control of multiple complex components. Besides, it enables nippy reconfiguration, which can help to implement other containers to the need of other experiments.

**Back-end:**
MONROE incorporates several repositories in the back-end, for maintenance and operation of the platform, for collecting metadata and results and temporarily store user data, and also to integrate and import mPlane data analytics computed on the traffic observed by the MONROE nodes. In particular, default experiment results and metadata are collected in a non-relational database using Apache Cassandra, which follows an experiment-oriented design.

**Visualization and open data:** Metadata and default experiment results can be freely visualized in near-real time. Specifically, MONROE offers a user interface that has been designed to provide a graphical representation of node deployment, their status, as well result collection of selected experiments. Measurement data will be openly released.

As shown in Fig. 3, users can fetch the data generated by their experiments from a temporary repository in the MONROE back-end. This is made available through the very same user interface that allows to schedule experiments.

---

4PC Engines, apu2 platform: https://www.pcengines.ch/apu2.htm
6https://github.com/MONROE-PROJECT
7https://www.monroe-project.eu/resources/papers/
8A MONROE node consists of several blocks, as described in what follows.
9User Access and Scheduling: It is not possible to login and run live experiments from a node. Instead, all measurements
10are managed by a central scheduling system that allows to deploy and control experiments over multiple nodes at the same time. Users access the scheduler through a user interface and public APIs. An AngularJS-based web-portal is available for such purpose, although users can also access the scheduler via a Command Line tool (CLI). In any case, platform users need to obtain a user certificate first, which is the same as per FIRE platforms and experimental federations that align with the Fed4FIRE initiative of the European Commission for a distributed, heterogeneous and large-scale experimental network platform. Indeed, as a consequence, concerning the scheduling APIs, the MONROE scheduling is compatible with other Fed4FIRE compatible interfaces.

**Maintenance and Management:**
This block is a subsystem used by the maintenance team to administer and maintain MONROE nodes with ease. It relies on a DB with a friendly interface to check node connectivity status and stats.

**Node:**
Measurement nodes are dual-apu2 machines with a Linux Debian based operating system. They include core components (e.g., for handling routing and connectivity, for monitoring the software behavior, etc.) plus a set of Linux Docker containers in which experiments are trigged in isolation. Moreover, some containers run constantly in the background to collect results from “default” experiments, e.g., to collect statistics on bandwidth throughput, delay, and gather metadata such as node GPS coordinates. The main idea behind using containers is that it allows for handy control of multiple complex components. Besides, it enables nippy reconfiguration, which can help to implement other containers to the need of other experiments.

**Back-end:**
MONROE incorporates several repositories in the back-end, for maintenance and operation of the platform, for collecting metadata and results and temporarily store user data, and also to integrate and import mPlane data analytics computed on the traffic observed by the MONROE nodes. In particular, default experiment results and metadata are collected in a non-relational database using Apache Cassandra, which follows an experiment-oriented design.

**Visualization and open data:** Metadata and default experiment results can be freely visualized in near-real time. Specifically, MONROE offers a user interface that has been designed to provide a graphical representation of node deployment, their status, as well result collection of selected experiments. Measurement data will be openly released.

As shown in Fig. 3, users can fetch the data generated by their experiments from a temporary repository in the MONROE back-end. This is made available through the very same user interface that allows to schedule experiments.

---

1http://www.fed4fire.eu/
2http://www.docker.com
3mPlane (http://www.ict-mplane.eu) is a measurement platform designed for fixed networks in the frame of a project funded by the European Commission, and which has been extended by MONROE to account for the analysis of traffic at mobile nodes.
4http://cassandra.apache.org
5http://visual.monroe-system.eu/
Remarkably, the measurement structure is not only capable of monitoring and analyzing the network behavior in real-time by scheduling user-defined Docker containers, but also to cache measurement and metadata for offline analysis.

**B. **MONROE Experiment Life-cycle

The workflow of MONROE experiments in depicted in Fig. 4. Initially, experimenters have to specify what kind of measurements they want to do and design them in Docker containers, that includes various software types. This is the design phase depicted in the figure, which includes creating and storing a Docker container in a github repository.

Secondly, containers have to be certified before going live. This is the central block in the figure. Specifically, containers have to be tested and, to this purpose, they can be run in special nodes devoted to debugging, to which users can connect directly and tune the experiments to respect safety and stability rules. In any case, a container has to be eventually certified by the MONROE maintenance team.

Finally, as shown in the rightmost part of the figure, the users deploy their containers into normal nodes through a user interface or the CLI to the scheduler that runs in the backend. The scheduler offers both Fed4FIRE-compatible and REST APIs, and connects to nodes to sense their availability and health conditions before making and enforcing any scheduling decision. The user interface, in addition to help to select types, number of nodes, and suitable time-slots, also helps users to collect passive and active results from multiple nodes. Indeed, after successfully finishing the experiments, experimenters can collect their data directly from the user interface along with metadata, e.g., GPS, throughput, and other pieces of background information on the nodes used in the experiment.

**IV. MEASUREMENT SETUP**

The measurement setup we have used consists in a Docker container with a WebRTC streamer and an IP tunnel handler that makes available a multimedia file over HTTPS. The container then includes a light implementation of WebRTC and we have made it publicly available for experimenters willing to design their own tests. When the experiment is scheduled on a set of machines, each of them makes a link available for connecting and watching the multimedia file using a Chrome browser acting as WebRTC client. We connect the WebRTC client from a computer in our lab to a MONROE node using such link. Therefore, the stream goes through the cellular uplink of the MONROE node, crosses the Internet and the access network of our lab, which uses a multi-gigabit connection. Hence, the bottleneck of the stream is in the cellular connectivity experienced by the MONROE node.

The MONROE platform provided us with many static and mobile nodes. Specifically, for the experiments reported in this paper, we have used the nodes listed in Table I. The table includes some limited yet important information about the location of the node. Indeed, we only report operator name and country, whereas more detailed pieces of information about the locations are omitted since this study does not serve as a thorough survey of operator’s performance comparison. The examples of results reported here are only meant to highlight how the potentials of WebRTC have not been fully unleashed by the sub-optimal MBB service available in many places in Europe as of today. However, as we will see later, MBB operators already offer WebRTC-ready connectivity for static users, while mobility is still a big issue.

The WebRTC throughput offered by MONROE nodes—i.e., the one offered by the MBB operator used at the node—differs a lot across space and operators, as well as across time, as shown in the third and fourth columns of Table I. Such throughput values have been collected with capacity experiment run a few minutes before and after running the WebRTC experiments.

To generate the figures reported later, we have used the data collected by Google Chrome on peer-to-peer connections, which include WebRTC streams. The resulting logs contain, per each individual stream, the timing and headers of packets received as well as the timing of various internal events such as received frames, losses, bitrate, delay and jitter.

**V. RESULTS AND OBSERVATIONS**

In this section we report statistics on WebRTC performance attributes, as observed by means of Chrome internals at the destination of the WebRTC streams. We run WebRTC streamers in static nodes across Italy, Spain and Sweden, and on mobile nodes in Sweden. We only report a small yet indicative subset of the results we have collected. For each experiment, we plot the following statistics: (i) BitRate, (ii) frames per second (FPS) rendered at the receiver, (iii) packet delay and (iv) jitter. We compare static and mobile cases.

**Static case.** In the first scenario that we consider, the streamer and the client have high quality connectivity (nodes 501, 591, and 596 in Table I). The selected MONROE nodes were connected to various 4G operators and, at the other end of the connection, we use a computer connected to a gigabit LAN directly connected to a multi-gigabit metropolitan network.

In most of the test cases, the media stream turned out to be smooth, the final user having a rather good experience.

### Table I: MONROE nodes throughput comparison

<table>
<thead>
<tr>
<th>Node ID</th>
<th>Location</th>
<th>Download [Mb/s]</th>
<th>Upload [Mb/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>423</td>
<td>Telia SE</td>
<td>0.178-10.38</td>
<td>0.165-1.58</td>
</tr>
<tr>
<td>428</td>
<td>Telia SE</td>
<td>0.53-13.97</td>
<td>0.25-2.4</td>
</tr>
<tr>
<td>429</td>
<td>Telia SE</td>
<td>0.78-12.178</td>
<td>0.22-1.74</td>
</tr>
<tr>
<td>501</td>
<td>Telia SE</td>
<td>0.34-0.26</td>
<td>0.22-0.3</td>
</tr>
<tr>
<td>591</td>
<td>Vodafone ES</td>
<td>24.58-72.64</td>
<td>17.56-23.14</td>
</tr>
<tr>
<td>596</td>
<td>Vodafone FI</td>
<td>32.78-65.41</td>
<td>14.21-25.10</td>
</tr>
</tbody>
</table>

13 docker.monroe-system.eu/deployed/monrtc
with typical bitrates of a few Mb/s, frame rates often above ten frames per second and latency in the ballpark of 100 ms, which is below the threshold of human perception. Jitter was normally below 100 ms. Fig. 5 illustrates an example of performance figures over time for three simultaneous connections using MONROE nodes in three countries. Of course each connection shows different performance figures, and in particular the Spanish sample shows the worst behavior in terms of rate and delay/jitter, although we need to mention that the selected MONROE node in Spain was using an Italian SIM card operating in roaming. In any case, the observed performance is acceptable.

**Mobile case.** In this scenario we select MONROE nodes in buses in Sweden (nodes 423, 428, and 429 in Table I), while the other end of the connection is in our lab, as in the other case. Fig. 6 shows that the user experience was not as smooth as in the first case since the bitrate was significantly degraded. Although WebRTC was able to keep the stream going, frame rate was very variable, and delay and jitter were annoyingly high in all locations and for all operators. We have observed many other cases like this in our measurements (and not only in Sweden, of course), which leads to the conclusion that current MBB networks are not ready to fully support WebRTC (and alike multimedia services) on the move.

To further understand the behavior of WebRTC on MBB networks, Fig. 7 shows 3D plots of bitrate and delay experienced by a MONROE node on a bus, for one of the operators in Sweden, as a function of the geographical position of the node (also shown in the topmost subfigure using Google Earth). The information about coordinates is provided by the MONROE node itself, within a metadata stream generated during the experiments. From the figure it is clear that some areas crossed by the bus were very poorly served, so that delay are constantly high and bitrates change smoothly over the trajectory, indicating that coverage (signal strength) is far from optimal in that area.
platform. To this aim, we have designed an open tool, running in a Docker container, for generating WebRTC sessions in mobile nodes by using standard WebRTC APIs. As such, the work presented a complete and novel methodology for the performance evaluation of web services using operational MBB networks. As an initial result, we have observed that mobility is still an important challenge for WebRTC, since MBB operators do not yet provide with full quality coverage for users on the move. Our approach can be used in the future for a broader and continuous assessment of WebRTC.

ACKNOWLEDGMENTS

Work funded by the EU H2020 research and innovation programme under grant agreement No. 644399 (MONROE). The work of V. Mancuso was supported by the Ramon y Cajal grant (ref: RYC-2014-16285) from the Spanish Ministry of Economy and Competitiveness.

REFERENCES


VI. CONCLUSIONS

In this work, we have evaluated the performance of WebRTC for static and mobile users by leveraging the MONROE