Experience: Implications of Roaming in Europe

(Article begins on next page)

01 September 2025

Experience: Implications of Roaming in Europe

Anna Maria Mandalari[1], Andra Lutu[2], Ana Custura[3], Ali Safari Khatouni[4],
Özgü Alay[5], Marcelo Bagnulo[1], Vaibhav Bajpai[6], Anna Brunstrom[7], Jörg Ott[6],
Marco Mellia[4], Gorry Fairhurst[3]

University Carlos III of Madrid[1] Telefonica Research[2] University of Aberdeen[3] Politecnico di Torino[4]
Simula Metropolitan[5] Technische Universität München[6]
Karlstad Universitet[7]

ABSTRACT

"Roam like Home" is the initiative of the European Commission (EC) to end the levy of extra charges when roaming within the European region. As a result, people are able to use data services more freely across Europe. However, the implications roaming solutions have on performance have not been carefully examined. 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 six countries across Europe. Using this platform, we measure different aspects of international roaming in 3G and 4G networks, including mobile network configuration, performance characteristics, and content discrimination. We find that operators adopt common approaches to implementing roaming, resulting in additional latency penalties of ~60 ms or more, depending on geographical distance. Considering content accessibility, roaming poses additional constraints that leads to only minimal deviations when accessing content in the original country. However, geographical restrictions in the visited country make the picture more complicated and less intuitive.

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 recently (as of June 2017) introduced a set of regulatory decisions [1] as part of the "Roam like Home" initiative. This initiative abolishes additional charges for users when they use voice and data
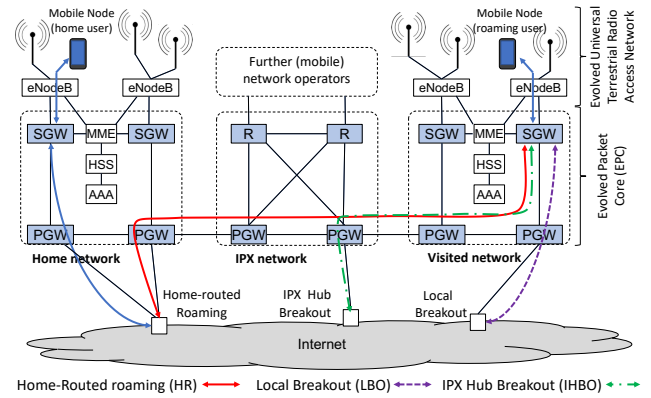
Figure 1: Internet access options for a mobile node at home (left) and when roaming (right).

services while roaming in EU. 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.

To support roaming, MNOs commonly connect with each other through an IP Packet Exchange (IPX) network. An IPX [2, 3] 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. In Fig. 1, we present a set of topology architectures that can be used for roaming in a mobile network, namely, home-routed roaming (HR), local breakout (LBO) and IPX hub breakout (IHBO). When a mobile node is at home (left, see Fig. 1), the *home user*'s traffic will take a short path inside the network to reach a suitable Packet Data Network Gateway (PGW) to the Internet. The traffic of a *roaming user* (right, see Fig. 1) is directed to an egress PGW whose location depends on the roaming architecture. In the case of HR, the mobile node receives the IP address from its home MNO and the *roaming user*'s traffic is first routed towards a PGW in the *home network* (red path). With LBO, the mobile node receives its IP address from the visited network
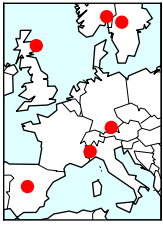
		MNOs	
NO	Telia NO	Telenor NO	
SE	Telia SE	Telenor SE	3 SE
UK	Vodafone UK	EE	
DE	Vodafone DE	T-Mobile	O2
ES	Vodafone ES	Movistar	Orange
IT	Vodafone IT	TIM	3 IT

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.

and the traffic is routed towards a local PGW in the *visited network* (purple path). When using IHBO, the mobile node obtains its IP address from the IPX networks and the traffic is routed through a PGW in the IPX network (green path). The topology can have a potential impact on the 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 is minimal when accessing services in the home country.

In this paper, we perform an extensive large-scale measurement study[1] 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?

To address these questions, we built 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 (see Fig. 2 and § 2 for details) across Europe, covering a total of 16 MNOs. We purchased Subscriber Identity Modules (SIMs) that support roaming for these MNOs and distribute them across the six countries. We characterize roaming operation and network performance (§ 3) and evaluate the impact on VoIP and web applications (§ 4) while roaming. We find that all observed MNOs use HR, which yields noticeable latency increases. We do not observe traffic differentiation policies for VoIP or web, but we do find evidence of content discrimination for roaming users. We review the existing work in § 6 and conclude the paper in § 7.

2 MONROE-ROAMING AND MEASUREMENT SETUP

In this section, we present the hardware platform we built for roaming measurements, and the manner in which we orchestrate it to collect our data.

Table 1: Terminology.

Home network	The network to which a mobile user subscribes.
Home SIM	The mobile user SIM while in the home country and connected to the home network.
Visited Network	The network to which a user connects while roaming internationally in the visited country.
Visited SIM	A user subscribed to visited network in the visited country.
Roaming User	A user subscribed to a home network in *country A* and who is roaming in a foreign *country B*.

2.1 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 Fig. 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 [8], 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 § 2.3), 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 a central server 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 [21].

2.2 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 Fig. 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

---

[1]The code and the dataset collected is open to the community: https://www.it.uc3m.es/amandala/roaming.html

[2]https://www.monroe-project.eu/access-monroe-platform/
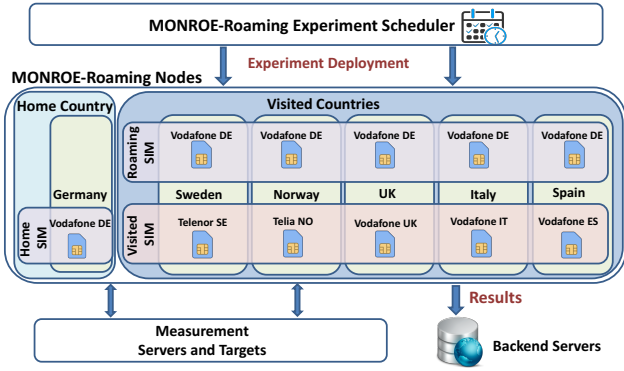
Figure 3: MONROE-Roaming platform and experimental setup. We exemplify our setup for Vodafone DE. We have five Vodafone DE SIMs in international roaming nodes and another SIM in the home country nodes. For each roaming Vodafone DE SIM, we insert the SIM corresponding to the local roaming partner for 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.

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 Fig. 3. We also describe the terminology in Table 1.

2.3 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 Fig. 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 § 3 and § 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 (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 in a timely manner. 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.

3 ROAMING SETUP AND PERFORMANCE
3.1 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 data transfers with popular URLs. We complement this analysis with metadata (e.g., radio access technology, signal strength parameters) collected from each node.

For each MNO, we measure in parallel the roaming user, the home user and the visited user (see § 2 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 details.

`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 data paths from each vantage point towards each measurement responder. Additionally, we collect the public mapped IP address for each vantage point (i.e., the IP

endpoint associated with the mobile client as seen from the public Internet).

dig: We run the `dig` utility for DNS lookups 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 inflate significantly performance metrics of web services (e.g., page load time), as well as impact the web experience of mobile users [14]. Thus, it is important to captute (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 to 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 2 times 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[3] 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 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.

3.2 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 data 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.

Next, we 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 measurement server, (iii) the number of home network PGWs that the roaming SIMs reach 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 five 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 for all the five 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). However, 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 explain or influence part of the differences observed between the MNOs.

Traceroutes, number of hops: We analyze our collected traceroute results from the roaming SIMs and compare 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 hops is the same.*[4] This is consistent with the HR configuration (Fig. 1), where the GTP tunnel 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 data 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, while for Vodafone UK we see that the same first hop appears on the data path, suggesting that the GTP tunnels of all our roaming users is terminated at a

---

[3]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.

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

Table 2: Distribution of the first IP interface and visited network per MNO. We report the number of networks each roaming user camps on in the visited country (# 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(%)).
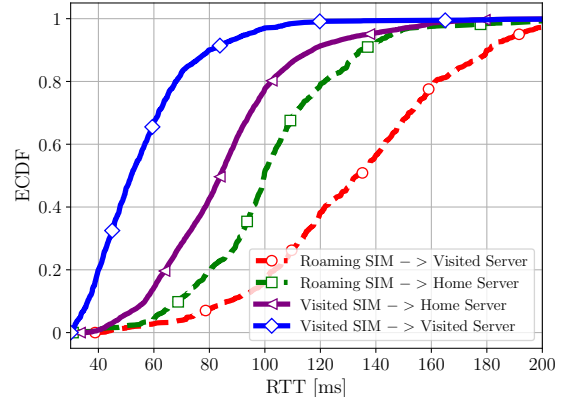
MNO	# of visited networks		# IP addr.	# tests	First hop breakdown(%)
	3G	4G			
O2 DE	9	9	20	657	1; 1; 1; 2; 2; 2; 2; 2; 3; 3; 3; 4; 4; 5; 6; 6; 9; 9; 11; 24
Telekom DE	5	5	4	1424	13; 19; 25; 43
Voda DE	5	6	2	1511	46; 54
Movistar ES	6	6	8	282	4; 5; 5; 7; 8; 21; 22; 28
Orange ES	7	7	3	900	6; 43; 51
Voda 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

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

single PGW. We also note that although for the majority of MNOs, these hops are configured with private address space, 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 data 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 [20]. We also check the time when the first IP address was used. We discover that all the PGWs are active in the same time. Multiple first IP addresses can be used at different time. We further contacted 3 MNOs and the information they provided about their network confirms our findings.

3.3 Home-Routed Roaming: Implications

Delay implications: The HR data implies that the roaming user's exit point to the Internet is always in the original home network (Fig. 1). Thus, the data that the roaming user consumes always flow through the home network. Depending on the location of the server, this translates to a potential delay penalty. Fig. 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). 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 best RTTs that one could expect with a LBO configuration, since LBO relies on access to local infrastructure with no need for tunnelling 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 very low penalties (e.g., just 0.2ms for O2 Germany).

We investigate this performance impact further and calculate the estimated delay penalty between LBO and HR when the target is in the visited network. 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. Fig. 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
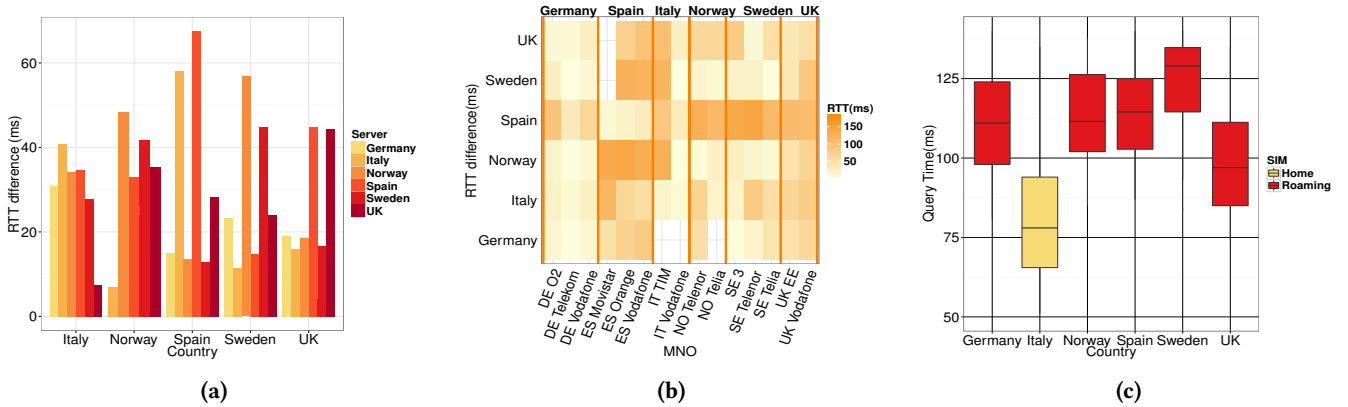
Figure 5: Delay penalty of HR: (a) RTT difference from the visited country to all servers for Vodafone DE; (b) RTT difference per operator; (c) DNS Query time to all FQDNs for TIM IT.

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 (and this is similar to the delay one would have because of the HR configuration).

We then evaluate the RTT difference between the roaming SIM and the visited SIM towards the same target and we group them per MNO. Fig. 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. Fig. 5c presents the distribution of DNS query times for all the SIMs of TIM IT. We note that for the home user the query time is significantly lower in average than for the other five roaming users. This is consistent for all the 16 MNOs we measured. This 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 be able to access the same content from a local cache (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 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.

4 VOIP & CONTENT DISCRIMINATION
4.1 VoIP Call

This section investigates potential traffic differentiation policies (such as blocking or throttling) that may hamper Voice over IP (VoIP) communications for a roaming user in comparison to a home user. We focus on three popular VoIP applications: FaceTime [4], Facebook Messenger, and Whatsapp [6].

Experiment Design. We begin with checking whether the MNO allows successful audio/video VoIP calls by carrying out some manual experiments with smart phone running the original applications and making calls among roaming SIMs. This lets us verify possible filtering or blocking in place at the time of the test. If successful, we then check for traffic differentiation mechanisms that could affect the call quality.

The experiment makes three audio and video calls using each application running on a regular mobile phone connected using an instrumented WLAN access point (AP) in our lab. Packet traces were recorded using `tcpdump` resulting in 18 traces, each with duration between [60,80] s. We verify the call-setup phase, which used a complex mix of TCP, STUN [25], and custom protocols to setup the end-to-end communication. From each trace, we then extract the actual audio/video streams. In the next step, we create a Docker container with pre-loaded traces, which we replay using `tcpreplay` to properly edit them so that packets are directed toward dedicated receivers hosted in our premises.

All applications run SRTP [10] on top of UDP, enabling easy adjustment of the packet timing and updating of the source and destination IP addresses. The dedicated server
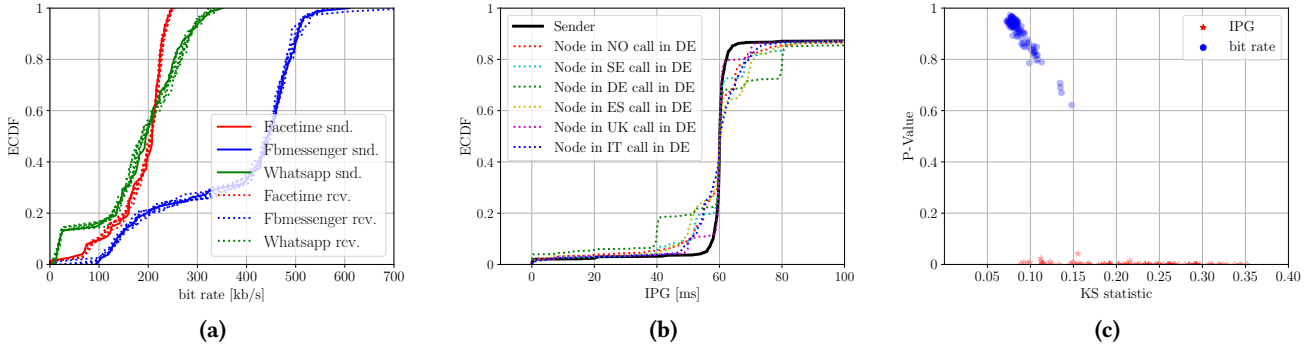
Figure 6: A sample of VoIP results and statistical similarity of bitrate and IPG: (a) Bitrate for operator O2 DE; (b) IPG for operator O2 DE; (c) KS and P-Value test.

in each country acts as a UDP receiver with a custom signalling TCP connection to log the status of the node (visited network, node identifier, metadata, experiment type, etc.) and the experiment associated with the receiver. In each test, the mobile node sequentially replays the pre-recorded traces with two receivers: a call to a destination in the home country, and a call to a destination in the visited country. For each call, we record pcap traces on both the client and server sides. We post-process these traces to check for traffic differentiation.

Results. We first verify that calls can be freely made from the tested network. We found that all operators allow users (even when roaming) to freely make VoIP calls using popular applications on their smart-phone. This confirms that no filtering was in place at the time of testing.

Next, we consider eventual traffic differentiation. We analyze well-known QoS metrics for real-time VoIP applications: packet loss, instantaneous bit rate, and Inter-Packet Gap (IPG), the time difference between two consecutive packets, to detect traffic differentiation. The results show that the packet loss ratio is less than 1% in all experiments. We conclude that no operators introduce artificial packet loss during our tests. MNOs could do so if desired to reduce the quality of calls for these applications and enforce traffic policing.

For each trace, we compare the bit rate we observe at the sender side and at the receiver side. Fig. 6a presents this as an ECDF for the three applications. Solid/dashed lines indicate the sender/receiver side when calling a receiver in the home or visited country (operator O2 DE). The applications use different audio/video codec combinations with different bit rate requirements. We observe no differences when using the tested network and the home network.

Fig. 6b shows a mobile user of the O2 DE operator making a Facetime call to our server in Germany while roaming. The periodic 60 ms long IPG is typical of low rate audio codecs the modern VoIP applications use. We observe some

differences when comparing measurements at the sender (solid line) and the receiver (dashed lines – one for each visited country). Some gaps are compressed (a smaller IPG), while others become expanded (a larger IPG). We observe this in all experiments with all operators when the sender is in its home country. We ascribe this to the modulation of the IPG by 3G/4G access mechanisms. Given the IPG is bounded to less than 80 ms, we conclude that this would not hamper voice quality, and expect these variations to be absorbed by the receiver playout buffer [11].

All experiments give very similar results, pointing to no evidence of traffic manipulation. We summarize these findings using the well-known KS Test [7] and *P*-Value [28] to determine whether the ECDFs differ between the sender (our reference) and the receiver. If statistically similar, the KS would be close to 0 while *P*-Value would be close to 1. If significantly different, the KS would be greater than 0, and the *P*-Value close to 0. Fig. 6c shows the scatter plot of the (KS, *P*-Value) points. Our results confirm that the receiver throughput is statistically identical to the sender throughput in all experiments. The IPG statistics are affected by the 3G/4G access mechanisms that alter the distribution (albeit not impairing the VoIP quality).

Finally, while QoS in terms of IPG and throughput are good, the total end-to-end delay could be significantly higher when roaming because of the HR solution. The one-way-delay could thus grow excessively large when two roaming SIMs call each other, making the interactive voice conversation difficult. The same effect was faced in GSM networks, and fixed by anti-tromboning [9] solution (e.g., allowing local breakout for voice traffic).

Takeaway. We do not observe any traffic differentiation on any of the 16 MNOs we measure. However, the additional delay of HR could impair real-time applications. This is an old issue (typically referred to as tromboning) which has been solved in GSM networks, but persists for 3G/4G VoIP applications.

4.2 Content Discrimination

In this section, we evaluate the availability of content when roaming, in particular whether MNOs filter website content and apply geographical restrictions. There are many reasons operators could have content filtering, which include complying with government guidelines or following court orders, e.g, to restrict access to file-sharing websites in the UK [24], or the use of 'opt-out' parental filters. We refer to any differences in the availability of websites and their content due to network interference as "content discrimination". When this difference is attributed to geographical location rather than the studied network, it is known as "content geo-restriction".

Experiment Design*.* The Open Observatory of Network Interference (OONI) [13] provides software tests for detecting censorship, surveillance and traffic manipulation in the Internet, using open source software. We ran two measurement campaigns using OONI's tool `ooniprobe` to detect network interference in home and roaming scenarios, geared at both content discrimination and content geo-restriction. We use the experimental setup described in Section 2.

The `ooniprobe` web connectivity test [5] performs the following steps over both the network of interest (tested network, using both home and roaming SIMs) and the Tor network [19]: resolver identification, DNS lookup, TCP connect and HTTP GET requests. First, ooniprobe performs DNS queries to disclose the IP endpoint of the DNS resolver in the tested network, and records the response, alongside a control response returned by Google's DNS resolver. Then, a TCP connection on port 80 (or port 443 for URLs that support TLS) is attempted using the list of IP endpoints identified in the DNS responses. HTTP GET requests are sent towards a list of URLs over both the tested network and over the Tor network and the responses are recorded. Differences in the results for the two tests are indicative of network interference. The results are made available to the public via the OONI API[5]. Our first set of tests considers 50 randomly selected websites from `ooniprobe`'s default global censorship list [12] contributed by users, seeking to identify content discrimination for roaming users. For the second set of measurements, we provide a list of 15 websites known to be available locally in the tested countries, but geo-restricted abroad. Beside network interference profile matching for the home and visited results, we additionally searched HTTP responses for known geo-restriction indicators and warnings. The high latency of the Tor network and data quota limitations limited us to test only a small number of websites.

Results*.* We create a network interference profile for each measurement, containing the names of blocked websites and

Table 3: Interference profile match Home vs. Roaming. High percentage indicate no difference between home and roaming setup.

	UK	NO	SE	DE	IT	ES
Vodafone UK	100%	98%	99%	98%	99%	86%
Telenor NO	99%	100%	98%	98%	98%	97%
Telia SE	99%	97%	100%	99%	99%	99%
Vodafone DE	98%	97%	99%	100%	98%	97%
TIM IT	98%	98%	98%	97%	100%	97%
Orange ES	99%	98%	98%	N/A	97%	100%

the HTTP body responses. For each visited website, we compare the responses observed using the plain connectivity against the one obtained using the TOR network. If those are different, this is an indication of manipulation. We then consider the home country interference profile as the baseline profile against which we compare the roaming profiles for that MNO to check if additional filtering is in place when roaming. If the HTTP response was unavailable due to censorship, we record the blocking mechanism as reported by ooniprobe.

We tested a mixture of dynamic and static websites and saw no content discrimination, with a perfect match of 95% and above of the tests between the home and roaming case. That is, if any censorship or blocking is present, it is the same in home and roaming case (this is consistent with HR again). The detailed results (shown in Table 3) are consistent with HR for all MNOs. The only exception is the case for Vodafone UK, where 3 websites were blocked (by DNS) in the home country, but were instead available in Spain. We double checked this, and found them to be a measurement artefact where ooniprobe incorrectly reported the blocking method, but we could not reproduce it later. We note N/A in the table where the results could not be collected at the time of measurement.

The findings are similar for the geo-restricted content tests. Here we check if a geo-restricted content that is accessible from home could be accessed also when roaming. To detect this, we search the body of the websites for known content indicating geo-restriction. We found results consistent with HR: content available in a user's home country remains available when roaming. Table 4 presents the profile match between the roaming profile and that of the visited network, for geo-restricted content. We did not observe substantial differences between the content policy of the two scenarios. We did also further (manual) investigation to verify video streaming application behaviour. We observed 10 cases where the website was accessible with no limitations, but then the actual video streaming was blocked with an alert to the user on the restriction.

It is worth noting that different countries may have different interference profiles. This can be due to, e.g., different court rules or legal guidelines. Given HR, a roaming user is

Table 4: Interference profile match for geo-restricted websites, Roaming vs. Visited. Differences are higher when comparing accessible content as home user or as roaming user.

	UK	NO	SE	DE	IT	ES
EE UK	100%	86%	89%	80%	87%	88%
Telenor NO	93	100%	94%	90%	88%	90 %
Telia SE	N/A	88%	100%	92%	90%	89%
O2 DE	87%	89%	82%	100%	N/A	84%
TIM IT	N/A	85%	83%	85%	100%	87%
Orange ES	N/A	85%	83%	N/A	89%	100%

always subject to his home country rules, even when visiting a foreign country where different laws are in place. This justifies the slightly lower figures in Table 4.

Takeaway. We found no evidence of content discrimination and geo-restriction for users in roaming scenarios, and the experience of browsing websites was the same in roaming and at home. However, there are clear differences between experience of a user of a network and a user visiting the same network, including the inability to retrieve geo-restricted content and the availability of different content.

5 DISCUSSION

5.1 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 an unprecedented 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 refer to just one snapshot – and the community should repeat these experiments over time to identify and investigate changes. The high cost of mounting such an experimental study is a major restricting factor for the density of sampling geo-locations. 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 similar number of samples per MNO and country. Furthermore, using the same equipment type throughout the measurement platform and in all locations eliminates potential device bias we might observe in the measurement samples.

Our measurement study focuses mainly on network performance and content implications of the roaming solutions in Europe. We leave for future work the exploration of potential performance penalties (see Section 3.1) on actual end-user Quality of Experience (QoE).

5.2 On Roaming Configurations

LBO appears a natural choice for an IP-based service and could offer lower operational cost 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 the major problem with MNOs that need to have near real-time view of the customer traffic for accounting reasons.

Whereas Session Initiation Protocol (SIP) signalling 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 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 [16]. 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). And if a service fails, it is not obvious who is responsible for finding the fault and fixing this. 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. 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 see 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 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 performance of application in different ways.

6 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.* [26] has leveraged crowdsourced measurements and focused only on national roaming agreements between MNOs in France. The study does not provide any further evaluation in terms of performance or content implications. Using controlled measurements in the dedicated platform MONROE [8] enabled Michelinakis *et al.* [22] to 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.

There have been myriad recent studies focusing on mobile network characterization and performance. For instance, while Huang *et al.* [15] study LTE network characteristics in a cellular operator in the US, Safari *et al.* [18] 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.* [17] 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 influence on network performance. The authors of [23] present a mobile app and a mechanism for identifying traffic differentiation for arbitrary applications in the mobile networks. Ververis *et al.* [27] surveys content filtering for a mixture of broadband and cellular ISPs and finds a lack of transparency around the policies they implement as well as outdated and poorly implemented blacklists. In our work, we not only focus on network performance of roaming infrastructure, but also identify possible traffic differentiation for particular applications and content discrimination and geo-restriction for users in international roaming.

7 CONCLUSIONS

While roaming internationally, different network configuration options can affect 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 Content Delivery Network (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. However, at the same time, the roaming user is still able to access (in majority of cases) the 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. The subscribers will increasingly start to 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.

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; accessed 06-March-2018].

[2] GSM Association: Guidelines for IPX Provider Networks. https://www.gsma.com/newsroom/wp-content/uploads//IR.34-v13.0-1.pdf. [Online; accessed 06-March-2018].

[3] GSM Association: IPX White Paper. https://www.gsma.com/iot/wp-content/uploads/2012/03/ipxwp12.pdf. [Online; accessed 06-March-2018].

[4] iOS 11: iOS Security Guide. https://www.apple.com/business/docs/iOS_Security_Guide.pdf. [Online; accessed 06-March-2018].

[5] Ooni - ooni: Open observatory of network interference.

[6] WhatsApp Encryption Overview. https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf. [Online; accessed 06-March-2018].

[7] Kolmogorov–Smirnov Test. The Concise Encyclopedia of Statistics, pp. 283–287. https://doi.org/10.1007/978-0-387-32833-1_214.

[8] 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.

[9] Aoun, C. Identifying intra-realm calls and Avoiding media tromboning. Internet-Draft draft-aoun-midcom-intrarealmcalls-00, Internet Engineering Task Force, Feb. 2002. Work in Progress.

[10] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and Norrman, K. The Secure Real-time Transport Protocol (SRTP). RFC 3711 (Proposed Standard), Mar. 2004. https://dx.doi.org/10.17487/RFC3711.

[11] Birke, R., Mellia, M., Petracca, M., and Rossi, D. Experiences of voip traffic monitoring in a commercial ISP. International Journal of Network Management 20, 5 (2010), 339–359. https://doi.org/10.1002/nem.758.

[12] Citizenlab. citizenlab/test-lists, Mar 2018.

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

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

[15] 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.

[16] 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.

[17] 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.

[18] 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 Teletraffic Congress (ITC 29), 2017 29th International (2017), vol. 1, IEEE, pp. 169–177.

[19] Loesing, K., Murdoch, S. J., and Dingledine, R. A case study on measuring statistical data in the Tor anonymity network. In Proceedings of the Workshop on Ethics in Computer Security Research (WECSR 2010) (January 2010), LNCS, Springer.

[20] 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.

[21] 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.

[22] 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).

[23] 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.

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

[25] Rosenberg, J., Mahy, R., Matthews, P., and Wing, D. Session Traversal Utilities for NAT (STUN). RFC 5389, Oct. 2008. https://dx.doi.org/10.17487/RFC5389.

[26] 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.

[27] 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.

[28] Wasserstein, R. L., and Lazar, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. The American Statistician 70, 2 (2016), 129–133. https://doi.org/10.1080/00031305.2016.1154108.