Quantifying the Influence of Regulatory Instructions over the Detection of Network Neutrality Violations

Marcio B. de Carvalho*, Vitor A. Cunha†, Eduardo da Silva‡, Daniel Corujo†,
Joao P. Barraca†, Rui L. Aguiar†, Lisandro Z. Granville*
*Institute of Informatics – Federal University of Rio Grande do Sul – Porto Alegre, Brazil
†Instituto de Telecomunicações, Portugal
‡Department of Informatics – Catarinense Federal Institute – Araquari, Brazil
*{mbcarvalho,granville}@inf.ufrgs.br, †{vitorcunha,dcorujo,jpbarraca,ruilaa}@av.it.pt, ‡eduardo.silva@ifc.edu.br

***Abstract*—The state-of-the-art solutions for detection of Network Neutrality (NN) violations assume that all detectable Traffic Differentiations (TDs) are in fact NN violations. However, legislators and regulatory agencies state instructions that establish which TDs may be considered as violations (or are allowed), and in which conditions. We advocate that these instructions should be considered before signaling a detected TD as an NN violation. In this paper, we are concerned with quantifying how much these instructions influence the results achieved by state-of-the-art solutions. We analyzed the public dataset of TDs detected by Glasnost under the regulatory perspective. We found that in specific circumstances, up to 48% of detected TDs cannot be conclusively signaled as NN violations. Our findings point towards the need for additional considerations when designing solutions focusing on NN, and to weaker conclusions drawn by solutions that ignore the regulatory perspective of the Internet.**

I. Introduction

Network Neutrality (NN) is a principle that has multiple definitions spread across the academic literature [1]. These definitions differ mostly on what constitutes the proper equality level to consider a network as neutral. For instance, they range from those that state that all traffic packets should be treated equally [1] to those that allow justifiable Traffic Differentiation (TD) practices [2]. NN definitions may also be included in the instructions that regulate the activities on the Internet. These instructions are set by legislators and regulatory agencies whose acts are valid just within a geographical area named *jurisdiction*, which usually encompasses a state, a country, or a region. Therefore, each jurisdiction may have its NN definitions. Thus, both academic literature and regulatory instructions provide multiple and heterogeneous NN definitions.

We advocate that the regulatory instructions should be used as guidelines to build solutions designed to detect NN violations. In this case, an NN violation would be the traffic management practices that were prohibited by the legislators and regulatory agencies (instead of a violation of academic definitions). Being based on regulatory instructions, the solutions may provide legally actionable evidence to support customer complaints against Internet service providers, for deliberation in the competent judicial authority [3]. Also, the regulatory instructions tend to be more detailed because they are used to support the activities of the regulatory agencies. They detail which TD techniques are prohibited (or allowed) to be applied over traffic from which applications, protocols, or services, and in which situations. For instance, is the throttling of Peer-to-Peer (P2P) protocols allowed without restriction, confined to just peak hours, or is it forbidden? Is zero-rating a mobile application allowed when the competitors are still charged? Can Internet contents be blocked, and if so, is a court order required or can anything be blocked arbitrarily?

The fact that regulatory instructions are valid just within the jurisdiction of the legislators, or the respective regulatory agencies, leads to the scenario depicted in Fig. 1. Each jurisdiction (Jurisdiction 1, Jurisdiction 2, ..., Jurisdiction n) may have its own NN definitions (NN 1, NN 2, ..., NN n, respectively). Thus, an end-to-end network path between an end-user and an application server may traverse different jurisdictions with different NN definitions. Also, even when the traffic is confined within one jurisdiction, it should be evaluated accordingly to the regulatory instructions stated in that jurisdiction. However, state-of-the-art solutions for the detection of NN violations are only designed to detect certain types of TDs (*e.g.*, a particular type of throttling, blocking, or prioritization), signaling any detected TD as an NN violation, regardless of the regulatory instructions for that jurisdiction.

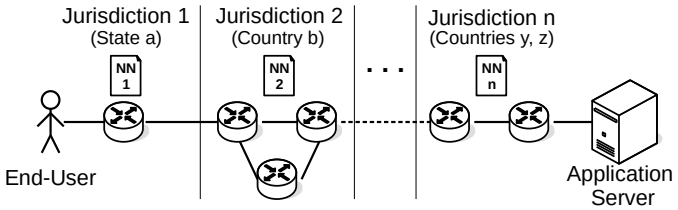


Fig. 1. End-to-end network path traversing multiple jurisdictions

We already raised concerns about the mismatch of regulatory instructions and the detection of NN violations [4], but we based that study on empirical evidence. In this paper, we aim to quantify the influence of using definitions from regulatory instructions on the detection of NN violations. Our analysis aims to answer the following questions: *(i)* How much different regulatory instructions agree whether a particular TD is considered as an NN violation? *(ii)* How many of the NN violations detected by state-of-the-art solutions

ISBN 978-3-903176-28-7

remain as violations considering the regulatory instructions? *(iii)* How much influence may the interpretation of regulatory instructions have over the results? *(iv)* How much influence the changes in the regulatory instructions have along the time?

To answer these questions, we analyzed the results publicized by state-of-the-art solutions for the detection of NN violations analyzing the detected TDs to verify whether they are, in fact, NN violations considering the corresponding regulatory instructions. We evaluated the public dataset of Glasnost [5] because it is the only one that provides the network captures and metadata that caused the TD verdicts. We opted to evaluate the data from 2016, which is the most recent complete year available in the dataset. Since it is the only dataset available, we designed three hypothetical scenarios to help answer the previous research questions.

The contributions of this paper are two-fold. *(i)* We are presenting evidence that solutions designed for NN must not ignore the regulatory perspective of the matter. This finding may seem obvious, but state-of-the-art solutions do not accomplish the problem in this way, taking the risk of drawing weak or wrong conclusions, as we show. *(ii)* We discuss modeling details found along this research that may help to design solutions for NN that take regulation into account.

This paper is organized as follows. In Section II, we present the state-of-the-art solutions to detect NN violations. In Section III, we present Glasnost dataset findings that are fundamental for the research conducted in this paper, including measurement results that were not explored by its authors. In Section IV, we present the chosen regulations that were used to evaluate the dataset. In Section V, we present our analysis and discuss its results. Then, we conclude the paper and provide the final remarks in Section VI.

II. Related Work

In this section, we present TD definitions that are related to aspects and definitions of NN. Then, we review the state-of-the-art solutions designed for the detection of NN violations. After, we present the solutions that already use NN definitions from the regulatory instructions. Finally, we present the previous work that raised jurisdictional issues in the NN context.

Traffic Differentiations (TDs) are characterized by their *Triggers*, *Traffic Classification*, *Differentiation Mechanism*, and *Perceived Discrimination* [1]. Triggers are properties of the flow (*e.g.*, application, path) or network conditions (*e.g.*, congestion) that leads an ISP to deploy the TD. Traffic Classification uses properties of the flow (*e.g.*, packet header or payload, behavior, routing) to determine its priority. Differentiation Mechanism is how the ISP interferes with the flow (*e.g.*, block, delay, drop, modify) to implement the TD. The Perceived Discrimination is the way users and detection solutions perceive the TD, which may be large delays, increased jitter, throttling, blocked traffic, or integrity violation.

There are multiple solutions designed for the detection of TD in the state-of-the-art, from which we present some representative ones. NANO [6] detects TD performing a confound factor analysis over data collected by multiple users. Several factors that may influence the performance of a flow are collected, such as those related to the client environment (*e.g.*, CPU and memory usage, OS), to the network (*e.g.*, IP, TCP state, TCP duplicates), and provided by the user (*e.g.*, ISP, Geo Location, SLA). If the confound analysis finds that the factor ISP is responsible for the performance degradation, then the ISP is performing TD. Diffprobe [7] adopts network tomography mechanisms, discovering internal network properties, to detect TDs. It detects which kind of management queue policy (Strict Priority, Weighted Fair Queue, Weighted Random Early Detection) is being used in the ISP routers. If it detects one of these queue policies, then the ISP is performing TD. NetPolice [8] is a solution designed to detect TD introduced by providers that operate at the core of the Internet. It is dependent on ISP network topology because it needs to target specific routers (*e.g.*, core routers) to measure packet losses (that usually happen when a TD is in place). If it detects packet losses, then the ISP is performing TD. Glasnost [9] detects throttling and blocking comparing tests of application and reference flows. The reference flow mimics the application flow (*e.g.*, BitTorrent, HTTP) differing its port or content to trick the Traffic Classification. When some discrepancy among these flows is found, the ISP is performing TD. These solutions do not adopt NN definitions from regulatory instructions. Thus, the detected TDs are considered NN violations without any support from the regulatory instructions.

Few solutions already adopt NN definitions from the regulatory instructions. The Body of European Regulators for Electronic Communications (BEREC) stated guidelines based on their regulatory instructions for deployment, by agencies of each European Union (EU) member, of solutions to the detection of NN violations [10]. However, the guidelines comprise just the NN definitions from BEREC, thus, ignoring other jurisdictions and the establishment of multiple definitions. Adkintun [3] is an infrastructure composed of a set of probes close to users to monitor the fulfillment of NN regulation in Chile. However, it only encompasses Chilean regulatory instructions, thus, also ignoring other jurisdictions and the establishment of multiple definitions. ISPANN [11] is a network auditor that inspects configurations of network devices to detect misconfigurations that violate NN definitions stated in the regulation. The network administrators choose, accordingly to their jurisdiction, the regulation that is used to check the configuration of their network devices. However, it is only meant as a confined configuration auditor; thus, it is not designed for end-to-end NN violation detection.

In a previous work [4], we already argued that the detection of NN violations should consider the diversity of NN definitions that may be found along an end-to-end Internet path. Therefore, such solutions should consider the jurisdictions and what is stated in their regulatory instructions to judge a detected TD as an NN violation. We proposed an architecture of a jurisdiction-aware NN violation detection adding an extra step (jurisdiction assessment) after current solutions' decision (TD detection and positioning). Therefore, the proposed prototype might be used alongside state-of-the-art solutions. However,

the architecture was designed based on several assumptions that were not validated in terms of reach or impact.

In this paper, we evaluate a public dataset of TDs detected by a state-of-the-art solution applying the NN definitions from regulatory instructions to judge them as NN violations. This analysis enables us to quantify how much these definitions will impact the results of such solutions. Although the analysis answer the research questions pointed in the previous section, we also aim to check assumptions made on the previous work helping to improve it. The findings of this evaluation allow the establishment of guidelines to academia to foster the development of solutions for the detection of NN violations that are in line with its regulatory aspects.

III. Glasnost Dataset

In this section, we analyze the Glasnost dataset presenting the aspects that are important for the evaluation conducted in Section V in which we quantify the influence of regulatory instructions over the detection of NN violations. As we focus on these aspects, most of the results presented here were not explored in the literature before.

Glasnost detects throttling and blocking. For these TDs, the traffic classification may be applied based on packet content (Deep Packet Inspection (DPI)) or application port usage (PORT). Throttling is detected based on the comparison of the performance of an application flow against reference flows. The reference flows are modified in their content (to bypass DPIs) or their ports (to bypass port-based classification). Blocking is detected when one or more of these flows fail to connect. Application and reference flows are tested in both directions (upload and download).

Glasnost was hosted on M-Lab from 2009 to 2017 [5]. The collected data is available along with the parser to build a dataset of detected TDs. To the best of our knowledge, Glasnost is the only solution that provides network captures of the detected TDs alongside all metadata information, which allows us to build a suitable dataset to research the detected TDs. We analyzed the dataset ranging from January to December of 2016, the most recent 12-months window with a stable amount of tests, which contains more than 2TiB of data. The dataset consists of test logs and dumps (PCAP format). The test log has performance information of the flows along with the test. The dump is kept as evidence when a TD is detected.

The parser [12] needs just the test logs to generate the dataset of detected TDs, that comprises 362GiB of logs for 2016. The parser has three main tasks, which we briefly present. *Import logs* reads the data from logs and imports them into the database. *Update geo&asn* annotates the database with the country and Autonomous System Number (ASN) of clients based on their IP addresses using GeoLite2 [13] and PyASN [14], respectively. Finally, *parse&analyze* annotates the database with the verdicts.

After the parser's processing, we have the verdicts of the 61773 Glasnost tests conducted during 2016. The parser judges the tests as "OK," "UNDEF," "OK1/2," "DPI," or "PORT." "OK" means that no evidence of a TD was detected. "UNDEF" can mean that the test was noisy or had the port changed during the test. "OK1/2" means that the test was "OK" in one direction (upload or download), but faced a problem in the other direction. "DPI" means that a TD (throttling or blocking) was detected based on the packet's content. "PORT" means that a TD was detected based on application ports.

In Figure 2, we show the respective percentage of verdicts of Glasnost tests along the year of 2016. We can see that those percentages remained steady throughout the year. The values are very close to the overall percentages for the year presented in Table I. The table also presents the conclusions we made over these verdicts: TD detected (DPI and PORT-based) at $\approx 20\%$, inconclusive tests (OK1/2 and UNDEF) at $\approx 35\%$, and no TD detected (OK) at $\approx 45\%$. Although not presented in Figure 2 and Table I, among the 12480 TDs detected by Glasnost in 2016, the proportion of blocking (2781, $\approx 22\%$) and throttling events (9699, $\approx 78\%$) is also relevant.

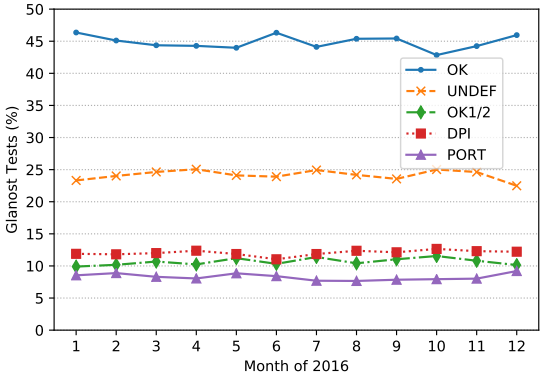
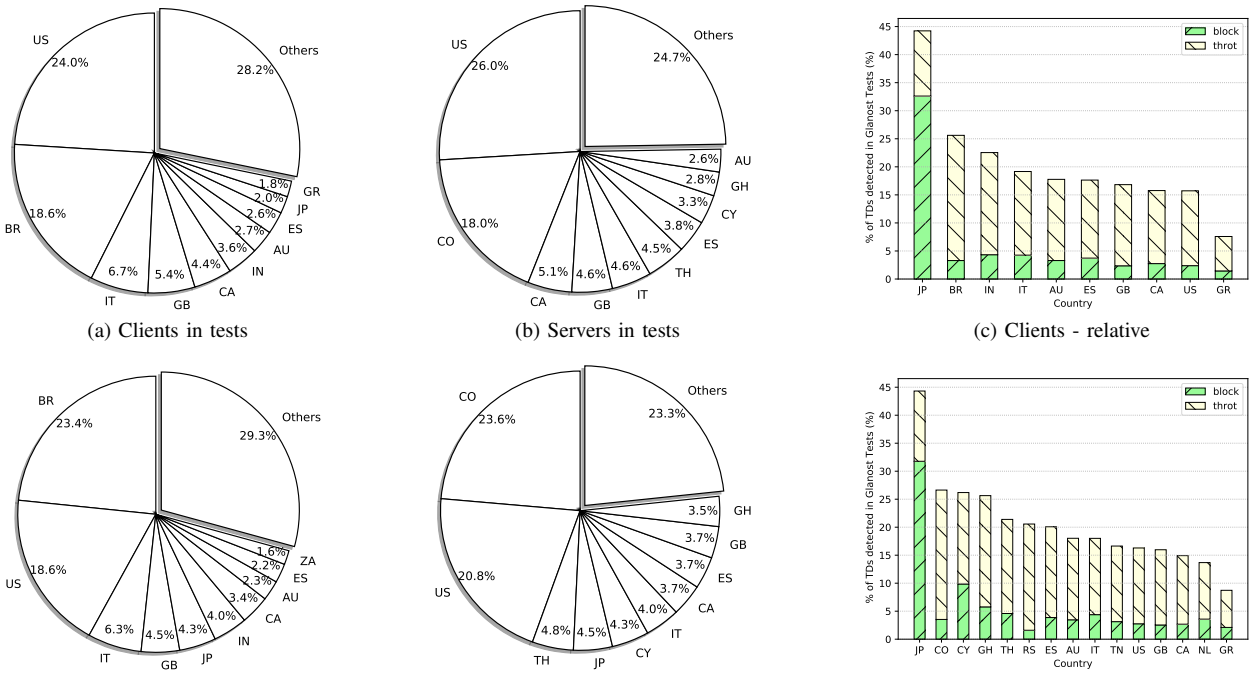


Fig. 2. Percentage of verdicts of Glasnost tests

TABLE I
Verdicts Numbers

Conclusion	Verdicts	# of Tests	% Verdicts	% Conclusion
No violation	OK	27954	45.2	45.2
Inconclusive	UNDEF	14823	24.0	34.6
	OK1/2	6520	10.6	
TD detected	DPI	7372	11.9	20.2
	PORT	5108	8.3	

The dataset comprises TDs that affected different applications. The proportion of each application within the dataset is different: BitTorrent (63.0%), FlashVideo (17.3%), HTTP (11.3%), NNTP (3.5%), eMule (1.8%), SSH (1.3%), POP3 (0.8%), IMAP (0.5%), and Gnutella (0.3%). These differences do not mean that one application is most differentiated against others. It is related to the application that users select to test in the Glasnost's interface (BitTorrent is the default test). As the Glasnost is a tool similar to a speed test, the chosen application is the only one that is tested, *i.e.*, there is no agent resident on the user's machine performing collection of other application's traffic. However, this skew in the proportion is important to take into account when we apply different interpretations of the regulatory instructions related to P2P traffic, as discussed in Section V, since they represent 65.1% of the tests.



(a) Clients in tests (b) Servers in tests (c) Clients - relative

(d) Clients with detected TDs (e) Servers with detected TDs (f) Servers - relative

Fig. 3. Distribution of countries with and without TDs and Distribution of countries with relative % of TDs over tests

The Glasnost parser performs the Geo Location of clients based on their IP addresses and ASNs. For each test, the dataset specifies the M-Lab server that is identified by the International Air Transport Association (IATA) code of the closest airport. Based on this characteristic, we constructed a table with the respective country of each M-Lab server using PyAirports [15], allowing us to determine the country of the client and server of each Glasnost test.

In Figures 3a and 3b, we present the top countries in the number of tests acting as clients or servers, respectively. The dataset comprises tests from 196 countries as clients and 29 as servers. We note that the distribution is slightly off between clients and servers. Brazil is the second country as a client, but it does not appear in the ranking as a server, because there was no M-Lab server hosted in Brazil in 2016. The tests from Brazil were likely routed to Colombia, which ranks second as a server but does not rank at the top of clients. The same applies to India that had their tests routed to Thailand. This discrepancy between the number of clients and servers exposes that a representative number of tests traverse country borders.

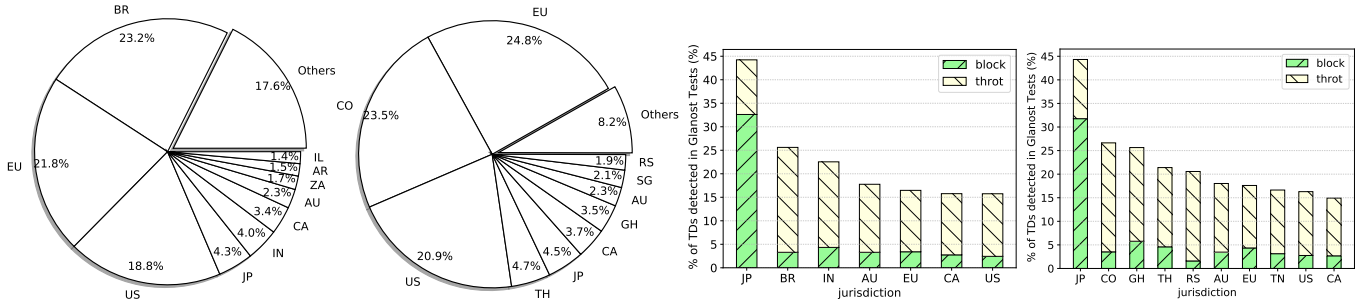
In Figures 3d and 3e, we present the distribution of countries in the number of detected TDs as clients and servers, respectively. Glasnost detected TDs in tests of clients of 165 countries and servers of 29 countries. Comparing Figures 3a and 3b against Figures 3d and 3e, we note that their distributions are very different. Brazil and Colombia became the first countries as clients and servers in TDs, respectively. Japan also climbs several positions in the distribution for TDs compared to the tests. This fact exposes that the proportion of tests that detected TDs is different in each country, which must be weighed against the absolute number of detections.

In Figures 3c and 3f, we present the top countries with the highest percentages of detected TDs as client and server, respectively. We filtered the results to present the countries that had more than 1000 tests to avoid noise because there are countries with few tests and a very high proportion of TDs. Japan is the first one in the proportion of TDs as clients and as servers. It is also interesting that most of them are blocking, contrasting with the overall proportion of blocking and throttling presented before (22% *vs.* 78%). Just Japan (in the presented charts) showed this behavior, exposing a massive appetite for blocking by Japanese ISPs. Few countries have proportions above the overall proportion of TDs ($\approx 20\%$): Japan, Brazil, and India as clients; and Japan, Colombia, Cyprus, Ghana, Thailand, Serbia, and Spain as servers.

In this section, we show that there is a representative number of tests that traverse country borders. Therefore, these tests need to be analyzed, considering the definitions stated in multiple regulatory instructions. We also show that the proportion of detected TDs and even the kinds of TDs deployed are different in each country. In the next section, we present the regulatory instructions from the top jurisdictions with detected TDs showing the heterogeneity of these definitions. Indeed, an analysis ignoring the existence of multiple and heterogeneous definitions may be inaccurate or even incorrect.

IV. NN DEFINITIONS FROM REGULATORY INSTRUCTIONS

In this section, we briefly present the regulatory instructions established in the jurisdictions that we considered in the evaluation conducted in Section V. As Glasnost detects throttling and blocking practices, we present the regulatory instructions that are specific to these kinds of TDs.



(a) Clients (b) Servers (c) Clients (d) Servers

Fig. 4. Top 10 jurisdictions with detected TDs

Figures 4a and 4b present the jurisdictional view of the data related to the tests that detected TDs, previously presented in Figures 3d and 3e without minding for jurisdiction. As the dataset provides the countries of the client and server of each test, we grouped countries into their jurisdictions, with clients and servers presented in Figures 4a and 4b, respectively. The top 10 jurisdictions for clients and servers are presented. We note differences from the data presented in Figures 3d and 3e. In Figures 4a and 4b, EU is the second jurisdiction for clients, and the first one for servers, because now it groups all tests related to EU countries and their territories. Despite accounting the TDs of the United States of America (USA) territories, the percentage of TDs remains unchanged due to their small contribution to the overall results (just 23 TDs). As we grouped countries and territories into their jurisdictions, other countries emerge in the Top 10 jurisdictions.

In Figures 4c and 4d, we present the top jurisdictions in the proportion of detected TDs, previously presented in Figures 3c and 3f without minding for jurisdiction. EU appears as the fifth jurisdiction for clients, and seventh for servers, because now it groups the results of all EU members, even those that do not appear in Figures 3c and 3f.

We consider the regulatory instructions from the Top 10 jurisdictions for servers and clients (Figures 4a and 4b) in the analysis conducted in Section V. As some jurisdictions appear in both Top 10s, we are considering regulatory instructions from 15 jurisdictions as detailed in Table II, which covers $\approx 81\%$ of TDs detected by Glasnost.

We researched the academic literature to find information about the regulatory instructions of each jurisdiction. When this research did not yield satisfactory results, we searched for information directly from the regulatory agencies responsible for the jurisdiction. We are able to understand the regulatory instructions written in English, Portuguese, and Spanish. For other languages, we translated the instructions using Google services. Due to this methodology, it is important to note that inconsistencies may be found. These inconsistencies may impose a small impact on the paper results. Despite this, our argument that heterogeneous definitions of NN exist across different regulatory instructions still applies. The point is that we need to consider different definitions, without mattering which particular instances. The work [16] referenced in a few cases, written in Portuguese, was an abundant source of references and may serve as an index to the different regulatory frameworks available worldwide. The regulatory instructions of each jurisdiction are presented next.

Some jurisdictions do not specify exceptions in their regulatory instructions or have not established regulation (which could include exceptions). Brazil (BR) is the only jurisdiction that prohibits throttling and blocking [17] without exceptional situations either. India (IN) established its former regulatory instructions focusing on Internet fees [18], thus allowing throttling and blocking without exceptions. For some countries, we found evidence of a lack of regulation, such as Australia (AU) [16], Serbia (RS) [19], and South Africa (ZA) [16]. For others, we have not found evidence either of the lack of regulation or its existence, such as Ghana (GH) and Thailand (TH), where we suppose that there are no regulations related to NN. Thus, TDs may be allowed in these countries including both throttling and blocking without exceptions.

Some jurisdictions prohibit throttling and blocking, while still allowing exceptions for both. Argentina (AR) includes exceptions for blocking, such as judicial claims or user requests (*e.g.*, parental control) [20]. Colombia (CO) includes exceptions for throttling (*e.g.*, congestion avoidance, security) and blocking (*e.g.*, prohibited or restrict use content, parental control) [21]. Israel (IL) includes exceptions that may be defined by the prime minister [22]. IN includes exceptions for throttling and blocking (*e.g.*, emergencies, restrictions on unlawful content, security, and integrity of the network) in their regulatory instructions stated in 2018 [23].

Some other jurisdictions prohibit throttling or blocking but allow exceptions only for one of them. EU establishes that throttling may be allowed under certain situations that the National Regulatory Authority (NRA) shall evaluate and decide [10]. The USA allowed blocking of content deemed illegal in its former NN regulation.

There are jurisdictions that allow throttling and blocking, as long as they can be considered justifiable, such as Canada (CA) [24] and the USA after 2018-06-11. Thus, providers are entitled to perform the traffic management practices deemed necessary, without explicit prohibitions established *a priori*.

Some jurisdictions prohibit one practice but allow the other. Japan (JP) prohibits throttling but allows exceptions such as for

TABLE II
NN DEFINITIONS WE CONSIDERED

Jurisdiction	NN regulatory instructions			
	begin validity	end validity	throttling	blocking
AR	2014-12-18	still valid	✗*	✗*
AU	–	–	✔	✔
BR	2014-04-23	still valid	✗	✗
EU	2015-06-30	2016-08-29	✗*	✗
	2016-08-30	still valid	✗*	✗
CA	2009-10-21	still valid	✔*	✔*
CO	2011-06-16	still valid	✗*	✗*
GH	–	–	✔	✔
IL	2014	still valid	✗*	✗*
IN	2016-02-08	2018-07-10	✔	✔
	2018-07-11	still valid	✗*	✗*
JP	2006	still valid	✗*	✔
RS	–	–	✔	✔
SG	2011-06-16	still valid	✔*	✗*
TH	–	–	✔	✔
US	2015-02-26	2018-06-10	✗	✗*
	2018-06-11	still valid	✔*	✔*
ZA	–	–	✔	✔

allowed (✔), prohibited (✗), has exceptions (*)

EU jurisdiction = AT, AW, AX, BE, BG, BM, CW, CY, CZ, DE, DK, EE, ES, FI, FO, FR, GB, GF, GI, GL, GP, GR, HR, HU, IE, IT, KY, LT, LU, LV, MQ, MT, NC, NL, PF, PL, PT, RE, RO, SE, SI, SK, SX, TC, VG, YT
US jurisdiction = GU, PR, US, VI

P2P or high demand users [25]. We have not found evidence of the prohibition of blocking in Japan, which may explain the already mentioned massive appetite for blocking by Japanese ISPs. Singapore (SG) allows throttling as long as justifiable (reasonable traffic management), but prohibits the blocking of legal content [26].

We need to point general comments about the regulatory instructions we discussed. Despite the more restrictive instructions in the EU regulation established in 2016 (compared to 2015), we conclude that there is no significant difference about throttling and blocking between both regulations. Japan has NN regulatory instructions established since 2006. We found evidence that this regulation remains the same since 2006 because there are guidelines, discussed in 2018, for the establishment of new rules [27]. The USA faced a hot debate around NN. For the sake of simplicity, we decided to consider just the two major eras in the US NN regulation, the Federal Communications Commission (FCC) regulation era, and the current lack of regulatory instructions era, ignoring the legal battles that suspended or re-established each regulation in specific periods [16].

After inspecting these regulatory instructions, we can note that there is a wide range of regulatory instructions established around the world differing about their allowances, prohibitions, and exceptional situations. This finding confirms our concerns about the NN violation detection based on regulatory definitions that motivated this paper, as presented in Section I.

V. RESULTS AND DISCUSSION

In this section, we present our analysis of the Glasnost dataset introducing NN definitions from the regulatory instructions. We quantify how these instructions influence the verdicts given to TDs signaled as NN violations due to the lack of regulatory interpretation. All results were gathered through Structured Query Language (SQL) queries in a PostgreSQL database, which holds the parsed Glasnost dataset.

A. Introducing Regulatory Instructions into Glasnost's Results

First, we analyzed the number of detected TDs whose clients and servers were in the same jurisdiction against those that were across jurisdictions. As for this, we do not need the NN definitions, just the jurisdictional area, we used all TDs in the dataset, instead of reducing the analysis to those that we got the regulatory instructions (Table II). We found that in 49% of detected TDs clients and servers are in the same jurisdiction, while conversely, in 51% of detected TDs they are in distinct jurisdictions. These values demonstrate a clear split between the detected TDs happening in a single or multiple jurisdictions. Despite the M-Lab infrastructure having broad global coverage and service that routes tests against the server closer to the client, we still have half of the TDs being detected with the client in one jurisdiction and the server in another. We assume that on the Internet, where the scale of Content Delivery Networks (CDNs) is much bigger, this proportion decreases, but we still may find a representative quantity of traffic traversing multiple jurisdictions, thus facing multiple NN definitions, as depicted in Figure 1.

Then, we analyzed the TDs introducing the regulatory instructions presented in Section IV. We restricted the analysis to the TDs whose jurisdictions we had got the regulations (Table II), thus reducing our dataset from 12480 to 10048 TDs ($\approx 81\%$). We evaluated four scenarios, as follows. As the Glasnost dataset comprises tests of 2016, we considered the instructions enforced at the time of the tests, which is the most realistic scenario. After this consideration, we introduced three hypothetical scenarios. One scenario is a time shift as if the tests were performed in 2019 (after the US NN regulation rollback and stricter NN rules in India). The other two hypothetical scenarios apply a stricter interpretation of the regulation (in 2016 and 2019) for those definitions that allow throttling and blocking of unlawful content. In this sense, we interpreted all P2P tests (BitTorrent, eMule, and Gnutella) as being related to unlawful content. We do not advocate against or for this interpretation; we want to analyze its impact on the results. Finally, we named these scenarios as 2016, 2019, 2016 (strict), and 2019 (strict), respectively.

As we may have different NN definitions stated for client and server jurisdictions, we investigated how much these definitions agree about a specific TD. In other words, it does not matter if each jurisdiction judges the TD as a violation or not, but whether both reach the same verdict. We considered only the TDs where divergences may appear, *i.e.*, when clients and servers are in different jurisdictions, which corresponds to $\approx 42\%$ of TDs. It is important to remember that we restricted the TDs to those that we have the instructions of the client and server jurisdictions listed in Table II. Thus, this percentage is slightly different from the presented before (49%), considering all detected TDs. We present the results in Figure 5.

We can see in Figure 5 that there is a significant number of agreements of different regulations. The top three jurisdictions drive the vast source of agreements for clients (BR, EU, and the USA) and servers (EU, CO, and the USA) that agree about

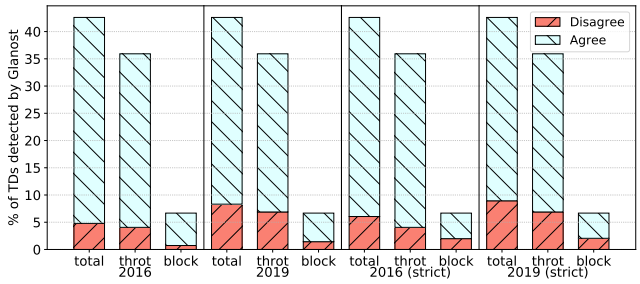


Fig. 5. Agreements and disagreements of regulatory instructions

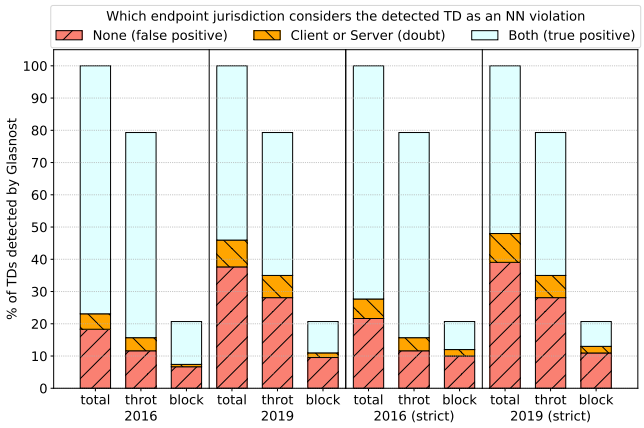


Fig. 6. Influence of regulatory instructions in verdicts

throttling and blocking practices in 2016. In 2019, we can see a rise in disagreements due to the NN regulation rollback in the USA. Adding the interpretation of P2P as unlawful content also increases the number of disagreements, due to the lack of this exception in Brazil (throttling and blocking) and Colombia (throttling). As P2P is involved in a massive part of TDs detected by Glasnost (65.1%), this interpretation impacts the results, especially in blocking because regulatory instructions mostly allow blocking of this content. As the content is deemed unlawful, it does not make sense to allow throttling (slow down) the access to the content.

The amount of disagreements presented in Figure 5 is small but representative, exposing the regulatory discrepancies and how they impact the results. These results show that one cannot judge whether the TD is a violation just with the information we have. We need more information to decide in which jurisdiction the TD occurred so that we can judge the TD under the right regulatory instructions.

We also analyzed the influence of the regulatory instructions when we judge TDs. Thus, we are interested in how much of the TDs detected by Glasnost cannot be considered NN violations under the regulatory perspective. We present the results in Figure 6, where we show in which jurisdictions (client or server) the detected TD is considered an NN violation. The legend "None" means that the TD is not considered a violation in client or server jurisdictions, thus it is a "false positive." They are indeed TDs, but they were interpreted implicitly as NN violations, which is wrong accordingly to the regulation of the endpoints. The remaining legends ("Client or Server," and "Both") are self-explanatory.

In 2016, we can see that $\approx$ 18% of TDs detected by Glasnost were false positives. Additionally, $\approx$ 5% of TDs are not considered violations on the client or server jurisdiction of the TD, thus, may or may not be false positives. Therefore, introducing the regulatory perspective, we may find that from 18% to 23% (adding the 5% that might be) of TDs detected by Glasnost could not be considered NN violations. The proportions are slightly different for throttling and blocking, having more presence of false positives for blocking ($\approx$ 32%).

In 2019, we can see a steep rise in the number of false positives ($\approx$ 37%). The TDs that are not considered violations on server and clients jurisdictions also increased up to $\approx$ 8%. Thus, after the USA NN regulation rollback, we may find that from 37% to 45% of TDs detected by Glasnost could not be considered NN violations. As we saw in 2016, the TDs related to blocking also present more false positives ($\approx$ 46%).

Adding the interpretation of P2P being the avenue for transiting illegal content, we also see an increase in results for 2016 (strict) and 2019 (strict). That is so because we are introducing more situations where a TD is allowed. In 2016 (strict), the number of false positives is $\approx$ 21% (compared to $\approx$ 18% observed in 2016). The situation that the TD is not considered a violation in client or server jurisdiction also increases to $\approx$ 6% (compared to $\approx$ 5% observed in 2016). The overall result is that we may find that from 21% to 27% of TDs detected by Glasnost could not be considered NN violations when we add this interpretation.

Adding the interpretation above to the 2019 results, we also see an increase in some values. In 2019 (strict), the number of false positives is $\approx$ 39% (compared to $\approx$ 37% observed in 2019). The situation that the TD is not considered a violation in client or server jurisdiction also increases to $\approx$ 9% (compared to $\approx$ 8% observed in 2019). The overall result is that we may find that from 39% to 48% of detected TDs could not be considered NN violations. It is important to note that for blocking, the proportion is even higher, achieving $\approx$ 52% for false positives and $\approx$ 10% for possible (on client or server endpoint) false positive, thus, achieving a surprising range of 52% to 62% of false positives related to blocking.

In Figure 7, we present a similar evaluation but just considering the USA jurisdiction to quantify how the NN regulation rollback that happened in 2018 influences the results. Its legends follow the same pattern of Figure 6. The Figure shows that the influence is clear comparing the scenarios of 2016 (2016 and 2016 (strict)) with the scenarios of 2019 (2019 and 2019 (strict)). Indeed, the number of confirmed NN violations (Both) exchanges with the number of false positives (None). The number of doubts (Client and Server) decreases because, in a reasonable amount of tests, the clients from other countries were targeting servers in the USA. As the TDs somehow related to USA (client or server) represents $\approx$ 25% of the TDs presented in Figure 6, we can note the representative influence of the USA NN regulation rollback in the overall results.

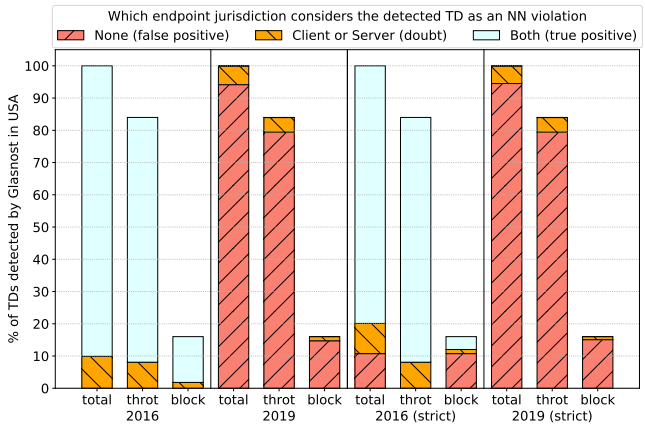


Fig. 7. Influence of NN regulation rollback in the USA

B. Discussion

We faced difficulty in finding a reliable source of information about the NN definitions stated in each jurisdiction. We inspected the raw regulatory instructions in some cases, but they are hard to find and be sure that it is the most up to date version; additionally, they are usually written in a foreign language. The access to this information should be easier for foreign researchers. It would also help to build systems based on reliable and easy to achieve regulatory information.

We perceived a shade of possible interpretations of detected TDs as violations. On one edge, we have the absence of NN regulation that all TDs are allowed and, therefore, there is no NN violation. On the other edge, we have the presence of a strict NN regulation that all TDs are prohibited and, therefore, all of them are NN violations. What drives the shade between these edges are the exceptions introduced in the regulatory instructions. The state-of-the-art solutions for NN violation detection assume a strict NN regulation that all TDs are classified as NN violations. Our analysis looks into different shades of gray, moving across time (after the USA regulation rollback) and adding more exceptional situations (P2P as unlawful content). As the regulation is alive along the time (see regulatory changes in Table II), the solutions designed for NN violation detection also need to consider these changes over time and the exceptions that it may introduce.

It is important to note that the presented results arise from an optimistic analysis of the Glasnost dataset under the regulatory perspective. As the dataset provides information only about the client and server involved in the test, we can only evaluate the jurisdiction of the endpoints. Thus, we can not assess the effects of any hidden jurisdictions along the network path of a test. While the client and server may be hosted in the same jurisdiction, the network path that connects them during the test may still traverse other jurisdictions, with different legal allowances than initially predicted. For instance, when considering a test with both the client and the server hosted in the USA, we may have the packets of this test going at some point through a router in Canada or even Mexico. The same may happen for clients and servers that are already in different jurisdictions. For instance, a client in Brazil performing a test against a server in Colombia may be routed through routers in Peru or Bolivia. The introduction of these hidden jurisdictions in the analysis could increase the number of disagreements and false positives, thus impacting the results we found.

The current USA NN regulation is also another source of inconsistent definitions, which may end in more disagreements and false positives. In our analysis, we considered that the regulation instructions are stated by countries or regions that define the jurisdictions. However, after the USA NN regulation rollback, some USA states established their own NN regulation, due to a lack of Federal regulation. Future work in this context should account for this situation in the USA.

To help to overcome the two points above, the modeling of solutions based on regulatory instructions should consider more complex traffic definition scenarios where the TDs take place. The models should allow the representation of the whole path between the client and application server that could expose these hidden jurisdictions (even smaller jurisdictions like states). However, as we know, the traffic between the client and the application server can follow multiple paths, thus, possibly transversing even more jurisdictions. Even worst, as we know, paths are dynamic and may be valid only for hours. Therefore, such models should allow the representation of the whole topology that was traversed by the application packets involved in a TD. Beyond this, solutions designed for detection of NN violations should have better precision when pointing where the TD is happening (at least to a country or state level) to help to establish the right jurisdiction. There are challenges to be faced in this positioning, as discussed in [28].

VI. Conclusion

In this paper, we have quantified the influence of regulatory instructions over the results of NN violation detection systems to confirm assumptions made on previous work [4]. Using NN definitions from the regulation, we analyzed the results available in the Glasnost dataset about TDs (throttling and blocking) detected in the year of 2016 hosted on M-Lab to answer our research questions. As these NN definitions are valid just within their jurisdictions, we used the Geo Localization information of clients and M-Lab servers to apply the correct definitions over the TDs detected by Glasnost. The answers to our research questions are as follows.

(i) How much different regulatory instructions agree whether a particular TD is considered as an NN violation? We found that the regulatory instructions of the endpoints agree (is a violation or not) about the verdicts of 91% to 95% of the detected TDs. This convergence is driven by jurisdictions that represent a vast amount of the detected TDs agreeing about blocking and throttling practices. However, the number of disagreements (5% to 9%) is not negligible and indicate situations we cannot judge the TD as a violation or not just with the information we have. It is needed to point where the detected TD occurred, at least at a country or state level.

(ii) How many of the NN violations detected by state-of-the-art solutions remain as violations considering the regulatory

instructions? We found that under certain circumstances, from 39% to 48% of the detected TDs are not NN violations accordingly to regulatory instructions. This result confirms our assumption that solutions must consider the regulation. Otherwise, their results have a good chance of being wrong.

(iii) How much influence may the interpretation of regulatory instructions have over the results? To answer this question, we interpreted the detected TDs related to P2P applications as being illegal content because some jurisdictions allow the blocking of such content. We found that from 16% to 22% of the detected blocking practices were no longer considered as NN violations. This finding exposes that solutions for NN that consider the regulation must model the normative interpretation somehow to accommodate this situation.

(iv) How much influence the changes in the regulatory instructions have along time? We focused our analysis on the regulation change that happened in the USA in 2018. The former regulation prohibited throttling and blocking but allowed the blocking of illegal content. The current regulation does not establish prohibitions *a priori*. Thus, all TD practices are permitted since they do not violate the antitrust principle. We found that the TD practices considered violations drop from 90% to 0%. We expected this huge change due to the paradigm shift among the former and current regulations. However, it exposes the need to model changes (especially the smaller ones that may be more probably) in the regulation that may happen and its validity over time.

Our findings expose that solutions for NN must consider regulatory instructions, or they need the support of another solution that is aware of these instructions, like JurisNN (presented in previous work). These findings seem to be obvious, but state-of-the-art solutions do not tackle them. Future work related to this paper includes the improvement of JurisNN to address the problems discussed in Section V-B, giving special attention to the need to expose the hidden jurisdictions that may be traversed in the path between the user and the server. After, JurisNN may be released to be used to judge whether TDs detected by state-of-the-art solutions are NN violations accordingly to the regulations, as performed in this paper.

ACKNOWLEDGMENT

We thank CNPq for the financial support. This research has been supported by call Universal 01/2016 (CNPq), project NFV Mentor process 423275/2016-0. This work is also funded by FCT/MCTES through national funds and when applicable co-funded by FEDER – PT2020 partnership agreement under the project UID/EEA/50008/2019.

REFERENCES

[1] T. Garrett, L. E. Setenareski, L. M. Peres, L. C. E. Bona, and E. P. Duarte, "Monitoring Network Neutrality: A Survey on Traffic Differentiation Detection," *IEEE Communications Surveys and Tutorials*, 2018.

[2] S. Jordan, "Four questions that determine whether traffic management is reasonable," in *2009 IFIP/IEEE International Symposium on Integrated Network Management, IM 2009*, 2009, pp. 137–140.

[3] J. Bustos-Jiménez and C. Fuenzalida, "All packets are equal, but some are more equal than others," in *Proceedings of the 8th Latin American Networking Conference, LANC 2014*, ser. LANC '14. New York, NY, USA: ACM, 2014, pp. 5:1—-5:8.

[4] M. B. de Carvalho, V. G. Schaurich, and L. Z. Granville, "Considering Jurisdiction When Assessing End-to-End Network Neutrality," *IEEE Internet Computing*, 2018.

[5] M-Lab, "The M-Lab Glasnost Data Set, 2016-01-01 up to 2017-01-01." [Online]. Available: https://measurementlab.net/tests/glasnost

[6] M. B. Tariq, M. Motiwala, N. Feamster, and M. Ammar, "Detecting General Network Neutrality Violations with Causal Inference," *CoNEXT*, pp. 289–300, 2009.

[7] P. Kanuparthy and C. Dovrolis, "DiffProbe: Detecting ISP service discrimination," in *Proceedings - IEEE INFOCOM*, 2010.

[8] Y. Zhang, Z. M. Mao, and M. Zhang, "Detecting traffic differentiation in backbone ISPs with NetPolice," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC*, 2009, pp. 103–115.

[9] M. Dischinger, M. Marcon, S. Guha, K. P. Gummadi, R. Mahajan, and S. Saroiu, "Glasnost : Enabling End Users to Detect Traffic Differentiation," *Proceedings of 7th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2010.

[10] BEREC - Body of European Regulators for Eletronic Communications, "BEREC Guidelines on the Implementation by National Regulators of European Net Neutrality Rules," 2016. [Online]. Available: https://tinyurl.com/y6hsf9dh

[11] V. G. Schaurich, M. B. de Carvalho, and L. Z. Granville, "ISPANN: A policy-based ISP Auditor for Network Neutrality violation detection," in *The 32-nd IEEE International Conference on Advanced Information Networking and Applications (AINA-2018)*, 2018.

[12] Max Planck Institute for Software Systems, "Glasnost," 2019. [Online]. Available: https://github.com/marcelscode/glasnost/

[13] Maxmind, "GeoLite2 Databases," 2019. [Online]. Available: https://dev.maxmind.com/geoip/geoip2/geolite2/

[14] Economics of Cybersecurity Research Group, Delft University of Technology, "PyASN," 2019. [Online]. Available: https://pypi.org/project/pyasn/

[15] Data61, "PyAirports," 2019. [Online]. Available: https://github.com/NICTA/pyairports

[16] L. E. Setenareski, "Fiscalização da neutralidade da rede e seu impacto na evolução da internet," Ph.D. dissertation, Federal University of Paraná, Brazil, 2017.

[17] Presidência da República, "Lei Nº 12.965, de 23 de abril de 2014," 2014. [Online]. Available: https://tinyurl.com/q5v43yc

[18] Ministry of Communications, "Telecom regulatory authority of india issues 'prohibition of discriminatory tariffs for data services regulations, 2016'," 2016. [Online]. Available: http://pib.nic.in/newsite/PrintRelease.aspx?relid=136211

[19] BDK Advokati, "Zero rating vs net neutrality – a (still) uncertain future in the eu and serbia," Information & Communications Technologies (ICT), Newsletter, Tech. Rep., 2017. [Online]. Available: https://tinyurl.com/y59nbqql

[20] Senado y Cámara de Diputados de la Nación Argentina, "Ley 27.078," 2014. [Online]. Available: https://tinyurl.com/y2en9rnn

[21] Colômbia. Congreso Nacional, "Ley 1.450 de 2011: por lacual se expideelPlan Nacional de Desarrollo, 2010-2014," 2011. [Online]. Available: https://tinyurl.com/y2tsqa64

[22] Knesset, "Communications Law (Telecommunications and Broadcasts), 5742-1982 - art. 51c," 2014. [Online]. Available: https://www.nevo.co.il/law_html/Law01/032_002.htm#med15

[23] Telecom Regulatory Authority of India, "Recommendations on net neutrality," 2017. [Online]. Available: https://tinyurl.com/y5o7nxvv

[24] CRTC - Canadian Radio-Television and Telecommunications Commission, "Telecom Regulatory Policy CRTC 2009-657: review of the Internet traffic management practices of Internet service providers," 2009. [Online]. Available: https://tinyurl.com/y67bcxv6

[25] K. R. Carter, T. Watanabe, A. Peake, and J. S. Marcus, "A Comparison of Network Neutrality Approaches In: The U.S., Japan, and the European Union," *Ssrn*, pp. 1–30, 2010.

[26] Info-communications Development Authority of Singapore, "Decision Issued by the Info-communications Development Authority of Singapore - Net Neutrality," 2011. [Online]. Available: https://tinyurl.com/y29k6b72

[27] T. Jitsuzumi, "Zero-Rating and Net Neutrality in the Mobile Market: The Case of Japan," *SSRN Electronic Journal*, no. 2015, pp. 1–24, 2018.

[28] Z. Zhang, O. Mara, and K. Argyraki, "Network neutrality inference," in *SIGCOMM 2014 - Proceedings of the 2014 ACM Conference on Special Interest Group on Data Communication*, 2014, pp. 63–74.