Journal of Network and Systems Management
https://doi.org/10.1007/s10922-020-09546-9

Towards a Holistic Net Neutrality Violation Detection System: A Case Study of Slovenia

Veljko Pejović[1]

Received: 5 January 2020 / Revised: 27 May 2020 / Accepted: 1 June 2020

Abstract
Democratic principles, from the freedom of speech, to fair business practices, rely on Net neutrality, i.e. equal access to communication infrastructure and services. While a number of national and international regulations stipulate Net neutrality, the actual enforcement is challenging as regulators have to collect and analyze a large amount of network measurements, and pinpoint cases of neutrality violations. Through a large-scale distributed crowdsourced measurements campaign, the Agency for Communication Networks and Services of the Republic of Slovenia (AKOS) has acquired a massive dataset of Internet performance measurements in Slovenia. In this work we analyze about one million multi-dimensional data records gathered by the AKOS Test Net measurement system and identify the practices, such as port blocking, that might violate Net neutrality principles. We then chart the limitations of the employed measurement approach and propose a holistic multi-stakeholder approach ensuring high quality measurement data upon which reliable Net neutrality violation inferences should be based.

Keywords Net neutrality · Mobile broadband networks · Network measurements · Data mining

1 Introduction

The Internet has transformed the way we work, communicate, socialize, and obtain information. It has thoroughly changed the way ideas are disseminated, news spread, and democratic movements organized. Connectivity is important for global economic development—everything else aside, access to information and communication technologies (ICTs) improves the GDP of a country by about 1% [1]. This is further exemplified as societies get more technologically advanced—in advanced

✉ Veljko Pejović
veljko.pejovic@fri.uni-lj.si

[1] Faculty of Computer and Information Science, University of Ljubljana, Večna Pot 113, 1000 Ljubljana, Slovenia

Published online: 06 June 2020

economies the Internet directly contributes to 21% of the GDP growth [2]. Furthermore, as the Internet of Things (IoT) becomes a reality, the Internet plays an increasingly important role in healthcare, transport, and factory automation [3].

The ever increasing importance of the Internet calls for the assurance that the global network represents, without discrimination, a level playing field for all the participants [4, 5]. While there is no common agreement on the specifics of this principle, termed *Net neutrality*, the following definition by the National Regulatory Agency (NRA) of India nicely summarizes the main postulates: "Net neutrality is generally construed to mean that [Internet service providers (ISPs)] must treat all internet traffic on an equal basis, no matter its type or origin of content or means used to transmit packets. All points in a network should be able to connect to all other points in the network and service providers should be able to deliver traffic from one point to another seamlessly, without any differentiation on speed, access or price. The principle simply means that all Internet traffic should be treated equally" [6].

Ensuring that the Internet is indeed *neutral* is one of the main challenges that NRAs nowadays face. Architected in the 1970s and backed by defense research funding, the Internet was conceived as a resilient decentralized network [7]. At the basic IP level, the Internet was designed to offer merely best effort packet delivery, no dedicated circuit connectivity, nor any Quality of Service (QoS) guarantees. In line with the main design guidelines behind the Internet and the fact that the initial network was fairly small, without prospects of it becoming the behemoth it is now, ensuring that the interconnected networks treat all packets/services equally (according to all possible definitions of equality) was not in the spotlight. Consequently, as noted by Crowcroft, the Internet is inherently biased [8]. For example, TCP flow capacity depends on the round-trip time from a sender to a receiver—the further the endpoints are, the lower the capacity is. Yet, as the Internet evolved, in particular on the higher protocol layers, and became a vessel for the World Wide Web (WWW) and a range of different services, such as Voice over IP (VoIP), the Net neutrality discussion rose to the level of application, i.e. service neutrality [9]. This is often apparent in regulatory documents, such as the EU 2015/2120 regulation stating that: "When providing internet access services, providers of those services should treat all traffic equally, without discrimination, restriction or interference, independently of its sender or receiver, content, application or service, or terminal equipment." [10]. It is in light of the above guidelines that we aim to understand how Net neutrality violation can be detected in practice.

In this work we collaborate with the National Regulatory Agency (NRA) of Slovenia—Agency for Communication Networks and Services of the Republic of Slovenia (AKOS)—and analyze the data AKOS has collected through crowdsourced mobile network measurements in Slovenia during a period of more than 2 years. The data contain measurements of download and upload transfer rates, ping round-trip times (RTT), as well as a set of QoS measurements, including port open/closed information, voice-over-IP (VoIP) traffic performance, and others, all measured strictly from a mobile client's point of view. The crowdsourced and uncontrolled nature of these measurements make the analysis challenging, as a number of entities outside of the measurement method's knowledge and control may lie on the path

from a client to a measurement server and impact the end-result of the measurements. Thus, in this paper we analyze country-wide measurement data with the goal of not only identifying potential instances of Net neutrality violation in Slovenia, but also outlining the limitations of the employed measurement method. We then critically discuss these limitations, propose various augmentations to the measurement methodology, and sketch a holistic method for Net neutrality violation detection. In our proposal we advocate for both crowdsourced measurements with user-controlled equipment (in order to provide a bird's eye view of the Internet) and measurements with controlled dedicated equipment (in order to zoom into behaviors of interest). Furthermore, we stress out the need for collaboration between Internet service providers (ISPs) and NRAs, as the measured performance can be objectively assessed only if the link subscription details are available. Finally, we discuss the data purity and statistical challenges, and propose machine learning-based methodology for in-depth data analysis.

2 Related Work

Owing to its open-for-interpretation definition, Net neutrality does not come with a clear set of guidelines that would ensure its protection. Instead, a number of NRAs have developed their own approaches to Net neutrality violation detection, and in this section, we present a brief overview of a few selected approaches. In addition, we discuss related work from the academic sphere, which often focuses on developing advanced methods for detecting particular strains of Net neutrality violation.

A recent study on Net neutrality regulation that Analysis Mason have conducted for the Body of European Regulators for Electronic Communications (BEREC) represents a good overview of the different methods that NRAs employ in order to monitor and enforce Net neutrality [6]. The study focuses on the solutions employed by NRAs in Chile, USA, and India and stresses the importance of a holistic approach to Net neutrality violation detection, the approach we also advocate in this paper. For example, the study examines Chile's NRA's requirement for national ISPs to provide detailed quarterly reports of traffic management practices and the achieved QoS and emphasizes a need for a *toolkit* rather than a single tool for network monitoring. Within the European Union individual NRAs have developed their own approaches under the umbrella of BEREC guidelines that provide a basis for Net neutrality violation detection in EU [11]. The German system relies on Breitbandmessung connection speed measurement app, which, among other information, asks users to state the details of the subscription package used.[1] The same approach is used by the Croatian NRA's HAKOMetar application [12]. The data analysis we conduct in this paper also emphasizes the importance of the information about the subscription over which a measurement is performed. Yet, aware of the limited knowledge a user might have about the actual subscription package (especially having in mind that terms of the contracts

[1] https://breitbandmessung.de/.

often change), we advocate an approach that relies on a direct dialogue between an NRA and an ISP. Another EU-based NRA, British Ofcom, relies on measurements with dedicated equipment developed by SamKnows [13]. After the limited explanation power of the crowdsourced measurements analyzed in this paper, we also call for measurements with calibrated dedicated equipment (see Sect. 5.3). However, we see NRA-controlled measurements as an augmentation, not a replacement for crowdsourced measurements, as these enable a larger scalability and a bird's-eye view of the networks [14]. Adkintun, the infrastructure developed in coordination with the Chilean National Secretary of Telecommunications (SUBTEL) for monitoring Net neutrality in Chile represents one of the most comprehensive network monitoring systems, as it relies on both specialized hardware (in form of Linksys WRT-160NL routers with a custom firmware) as well as on software probes (available for Linux, Mac, and Windows machines) [15]. Crowdsourced deployment enabled Adkintun to engage about 10,000 users nationwide and collect a sufficient amount of network speed measurements to result in a lawsuit related to ISPs failure to deliver the advertised speeds. Being one of the pioneering attempts, Adkintun was geared towards overall latency, jitter, and bandwidth measurements. In our work, we investigate a wider range of Net neutrality violations, such as port blocking, traffic shaping, and others. Furthermore, we go beyond data collection and simple descriptive statistics comparison and present a robust statistical methodology for violation detection.

In parallel to broader measurement initiatives lead by NRAs, academic researchers have developed detection methods for specific aspects of Net neutrality violations [16]. Glasnost detects BitTorrent traffic differentiation by comparing the performance of a pair of flows: a BitTorrent flow and a reference flow belonging to a different application [17]. NeutMon also concentrates on BitTorrent, and in a pan-European study shows traffic differentiation among a number of ISPs [18]. In a follow-up work, the authors show that besides throttling, BitTorrent traffic may be subject to discriminatory routing policies [19]. BitTorrent is just one of the applications that may be targeted by traffic shaping. NetPolice uses traceroute-like probes to cover a range of ingress/egress ISPs and source/destination pairs, and then performs a Kolmogorov–Smirnov test to identify differentiations over five Internet applications [20]. Similar to NetPolice is DiffProbe, a tool, that also detects differentiation by comparing two flows, yet addresses some of NetPolice's methodological deficiencies [21]. While Glasnost, NetPolice, DiffProbe, and a few other proposed approaches [22–24] rely on active measurements, NANO uses passive observations of the real network traffic [25]. The approach first stratifies the data according to the confounding variables (e.g. time of the day, Web browser type, etc.) that may impact the measurement results. Then, for each stratum, NANO estimates the performance change when a service is used via a particular ISP. Passive measurements bring two major benefits to Net neutrality violation detection. First, using the actual users' traffic, a passive measurement-based Net neutrality violation detection system is *stealth*, i.e. ISPs cannot recognize probing traffic and handle it separately from the actual client's traffic. Second, passive measurements do not incur any cost for a user—especially with mobile active measurements, one has to be careful about a potential data transfer cap a subscription might be tied to. Still, passive measurements depend on

the actual usage, may never shed light on the performance of less popular applications, and cannot saturate links in order to gauge connections' limits.

Network middleboxes can be used for various means of traffic shaping/differentiation, as summarized by Choffnes et al. in [26]. Here, the authors present evidence of middlebox-enabled policies that violate Net neutrality in production ISP networks in USA. The evidence has been gathered through studies that use a *record and replay* technique, where network traffic generated using an actual application, such as YouTube or Netflix, is recorded and replayed in both *exposed* and *hidden* (encrypted through a VPN) manner [27]. A comparison between the performance of flows sent via the former and the latter points out to potential differentiations by the ISP. The method has its drawbacks, however, as there are no guarantees that latter flow is treated differently simply because it uses a VPN (irrespective of the application), nor one can guarantee that network conditions remain the same for the whole duration of the two flows. Goel et al. propose a method for TCP flow splitting detection that is based on passive observations conducted in collaboration with content delivery networks (CDNs) [28]. The method analyzes timestamps of TCP handshakes and recognizes *fingerprints* of middlebox tampering with TCP/IP header bits. Despite only a handful of published approaches, the identification of middlebox-driven differentiation is crucial for Net neutrality violation detection. Results of the analysis presented in our work also emphasize the need to identify TCP connection splitting, media content transcoding, and other manipulations performed by middleboxes.

Despite the above NRA and academic efforts, Net neutrality violation detection is still fraught with issues.[2] The approach based on crowdsourced measurements analyzed in this paper is not an exception. For instance, the identification of potential confounding factors that may impact the measurement results, a problem acknowledged by Garrett et al., is an issue we faced early on in our project. We tackle this particular issue with the *enrichment analysis* elaborated in Sect. 6. However, this paper does not answer to all the known issues of Net neutrality violation detection. Rather, we conduct an analysis of country-wide crowdsourced measurement data with the goal of inferring Net neutrality violations and point out to the limitations of a single-faceted approach. Based on these results, we then propose a holistic method to address the remaining issues. For instance, a lack of ground truth on the expected connectivity performance remains a key problem in Net neutrality violation detection literature [16] and also in our work. Thus, in Sect. 5.4 we argue that ISPs need to be systematically involved in the measurement process by providing information about the examined subscriptions. Finally, to cope with the ambiguity of the sheer concept of Net neutrality, Garcez Schaurich et al. propose a system that takes as input country-specific Net neutrality rules and audits an ISP network, identifying potential violations [29]. The work aims to bridge the policy and the technical sides of the Net neutrality debates and provides a framework for rigid definition of key performance indicators (KPIs) that should be assessed in order to detect Net neutrality violations. As such, the work is orthogonal to our proposal (presented in

[2] Garrett et al. provide a comprehensive summary of issues in their recent survey of traffic differentiation detection [16].

Sect. 6) that follows from a large-scale data analysis. Namely, we propose a holistic measurement and analysis framework that ensures that the defined KPIs are indeed assessed in a statistically meaningful way.

3 Data and Methods

3.1 AKOS Test Net

The AKOS Test Net system was deployed in 2015 by the Agency for Communication Networks and Services of the Republic of Slovenia (AKOS) with the goal of providing a comprehensive view of Internet connectivity performance as experienced by end-users. Guided by the European Union BEREC directive, the system is envisioned as an essential tool for Net neutrality violation detection. The measurement system was designed by an independent company—Specure—which has developed similar solutions for the national regulators of Austria, Czech Republic, Slovakia, and Serbia, among other countries.

AKOS Test Net is composed of a mobile (Android and iOS) and a Web application, as well as the backend infrastructure. The main part of the backend is a test server to which a connection can be established and data transferred to/from. The applications enable parallel crowdsourced on-user-demand measurements of various connectivity parameters, including:

- Connectivity speed (upload/download transfer rate and ping RTT test);
- Connectivity parameters (connectivity type, received signal power, etc.);
- GPS location;
- QoS parameters (TCP/UDP port availability, sample URL page access, DNS request tampering, etc.);
- Network Diagnostic Tool (NDT) test results.

A breakdown of the parameters used in our analysis is shown in Table 1, whereas more details about the measurement process can be found in [30].

AKOS Test Net applications are publicly available[3] and not restricted to Slovenian users (albeit, a large majority of the results in indeed origin from Slovenia). A part of the measurement results is available as open data at the AKOS website. For the analysis performed in this paper, however, we obtained access to the complete measurement database, which, at the time of the analysis, contained around 950,000 individual speed test measurements[4] collected from June 2015 till December 2017.

[3] https://www.akostest.net/en.

[4] Each measurement session always includes speed test results. Other measurement types may or may not be executed depending on the application type (e.g. only the Android-based application supports TCP/UDP port availability test), client's permission (e.g. NDT tests need to be explicitly enabled), and sensor availability (e.g. GPS coordinates available or not).

Table 1 AKOS Test Net application measurement parameters used for Net neutrality violation detection analysis

Parameter	Description	Open data
Download and upload speed	The overall upload/download rate calculated in a 7-second test preceded by a pretest. In the pretest phase, the client opens three connection threads to the server and requests randomly generated packets. If the pretest shows that fewer than four packets have been successfully transmitted, the client keeps only one active thread. A full test is then executed over all three or just a single thread [30]	Yes
Ping RTT	The client sends pings via one of the opened TCP connections to the server in short intervals. The server responds with acknowledgements. The client measures the time between sending and receiving the return message, while the server additionally measures the time between sending its return message and the client's reception response. The client stores all measurements, and the median of all measurements is used as result	Yes
Network name	Name of the mobile network. Due to occasional glitches with descriptive names reported by mobile base stations (e.g. "Telco A" and "TelcoA" referring to the same provider), we first examine the network mobile country codes (MCC) and mobile network codes (MNC) and then select the most common network name as a label.	Yes
Connectivity technology	Technology category of the network, e.g. 3G, 4G, WLAN, etc.	Yes
Network type	Type of the network, e.g. GSM, EDGE, UMTS, HSPA, LTE, LAN, WLAN, etc.	Yes
Port blocking	The client sends TCP or UDP packets to the specified port and waits for a reply. The network quality test server receives these packets and responds to them. The control server compares the transmitted and received packets. It measures whether (TCP-/UDP-) packets have arrived or were blocked, how many packets have arrived or how many packets had to be resent. In addition, the variance of the arrival time can be calculated. This test can be configured to target different ports and to benchmark success against different timeout values (in the presented study a 3 s timeout was used)	No
VoIP blocking	The client sends UDP packets over a defined VoIP-port (5060) using a defined time interval. The server receives these packets, measures the time delays between them, jitter, and packet loss rate. The test is marked as successful if both incoming and outgoing packet jitter stays below 50 ms, and if no packets are lost	No
LTE RSRP	LTE signal strength in decibels	Yes
LTE RSRQ	LTE signal quality in decibels	No

3.2 Data Analytics Tools

We developed RICERCANDO mobile broadband measurement mining toolkit [31] driven by the postulates that network traffic analysis should include statistical analysis that goes beyond simple ad-hoc solutions, visualization and multidimensional exploration by networking experts, advanced machine learning modeling algorithms, and should allow the data to be pipelined to other tools [32]. RICERCANDO contains tools for big multidimensional measurement data transformation, geo- and time series-visualization, and advanced statistical analysis. RICERCANDO's visualization is designed to handle about one million data points per day. Our dataset is much sparser, thus we construct simpler custom solutions for data management and visualization, but rely on tools from RICERCANDO, namely the framework's *Significant Groups* Orange widget (see Sect. 4.3), for advanced statistical analysis.

The tools we develop and use are split in two groups:

- **Descriptive analysis tools** These tools, in the form of Jupyter Notebooks, were developed specifically for the AKOS test net analysis.[5] They enable connection with the AKOS Test Net PostgreSQL database, data processing with the help of numpy Python package, and matplotlib-based visualization.
- **Statistical analysis tools** These tools are a part of a general mobile broadband measurement mining framework and have been released as a separate open-source project RICERCANDO.[6] The tools rely on Orange, a popular data mining suite developed by the Faculty of Computer and Information Science, University of Ljubljana [33]. Combining Orange's wide range of data analysis widgets, and purposely-built widgets for uncovering factors that characterize the difference between two sub-datasets, we construct AKOS Test Net advanced statistics pipelines.

The descriptive analysis and the statistical analysis tools are connected with the help of pandas Python module. More specifically, data extracted from the database and processed in a Jupyter Notebook can be saved to local storage as a pandas DataFrame. An Orange widget we have developed allows these data to be imported in Orange workflows for further processing.

4 Data Analysis Results

The goal of our analysis is to evaluate the potential of distributed crowdsourced measurements for Net neutrality violation detection. In this work we adopt the EU 2015/2120 regulation's definition of Net neutrality (Sect. 1). However, the regulation provides merely a policymaker's level definition of Net neutrality, thus, we

[5] https://github.com/vpejovic/netnevt.
[6] https://github.com/ivek1312/ricercando.

define the following key performance indicators (KPIs) for detecting Net neutrality violations in the analyzed data:

- **Network data transfer speed differentiation:** under the same operating conditions, we hypothesize that there are no statistically significant differences among upload or download speeds measured between different groups (e.g. between users of different ISPs[7]). Otherwise, we assume that Net neutrality is violated;
- **TCP/UDP port blocking:** we hypothesize that within a single ISP there is no differentiation among services using different TCP/UDP ports, indicated by no statistical differences among the ratio of successfully transferred packets via different ports.
- **Application-level performance:** we hypothesize that relevant metrics characterizing an application-level service will not be significantly different between measurements taken within two different groups (e.g. users in different regions, connected to different ISPs, etc.). In particular, we focus on VoIP service and the following related metrics: inbound/outbound packet delivery rate and packet inter-arrival time jitter.

The selected KPIs are by no means an exhaustive set of Net neutrality indicators. However, the data we have at our disposal provides only end-user perspective—we do not have measurements originating from the ISPs or the network core. Indeed, we know nothing about the ISP policies, thus have no ground truth data about Net neutrality violations. Consequently, we base our inference on the comparison among aggregate end-user performance experienced with different operators and/or different services in similar contexts of use. Here, however, we have to emphasize that different ISPs have different infrastructures and peering agreements, to mention a few factors that remain outside of our knowledge yet may impact the performance measurements. Thus, with this analysis we also aim to uncover the limitations of an approach that relies solely on data collected by end-users.

4.1 Speed Test Analysis

The Net neutrality principle prohibits discrimination among different services and different users when it comes to download/upload transfer rates. This, however, does not preclude ISPs from offering a range of connection packages to users. These might come with different connection speeds, dynamic speed adaptation (e.g. higher rate in evenings and weekends), or data transfer amount caps, after which severe throttling might be in place (e.g. 10 GB at 10 Mbps, 128 kbps after that). Furthermore, Net neutrality regulations often allow temporary throttling, if such an action is

[7] While technically differentiation between users of two different ISPs does not necessarily amount to Net neutrality violation, we begin our analysis with this problem as it represents the easiest case—failing to detect differentiation at the ISP level indicates that differentiation on deeper levels of the hierarchy is unlikely to be detected either.

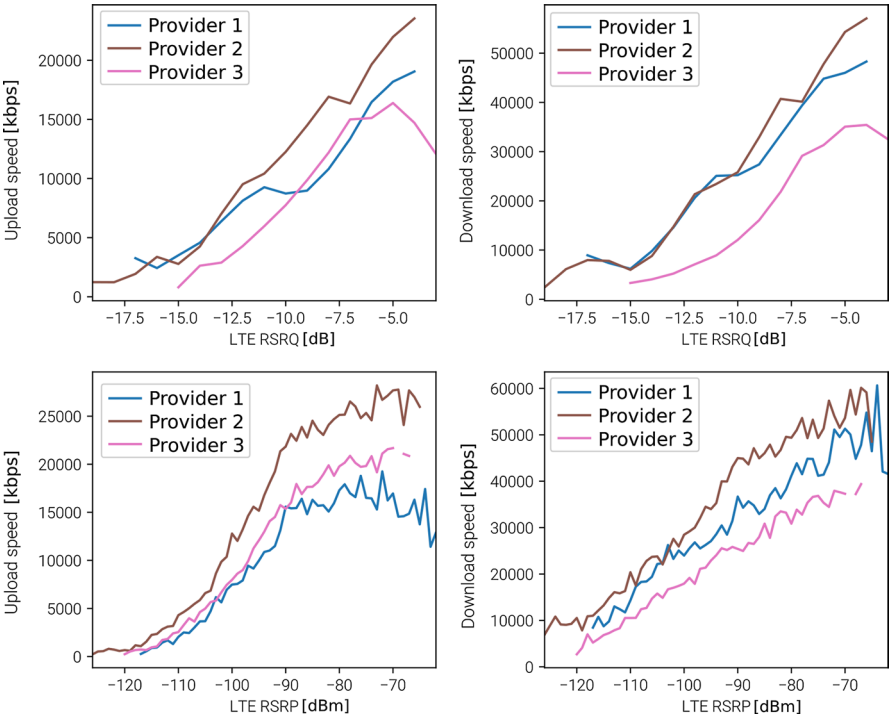

Fig. 1 Comparison of upload and download speeds of three major Slovenian ISPs. For a fairer comparison, the measurements are aligned over the common LTE RSRQ and LTE RSRP parameters

necessary for ensuring connectivity during occasional situations where the network infrastructure is overburdened (e.g. flash crowds).

AKOS Test Net data contains merely end-user speed measurements with no information on either the contract-prescribed speeds that a user is paying for, nor the information on operators' activities for ensuring network functioning in periods of high network strain. Moreover, conducted in the same manner and with the same network-level settings (TCP, always the same port), the measurements do not allow us to inspect potential throttling of different services. Thus, in our analysis we focus on comparing the achieved speeds across different operators.

We focus on LTE, the best performing connectivity technology in our dataset, since, if the throttling indeed happens, it will most likely be observable when the least restrictive technology is used. Furthermore, LTE provides a set of metrics enabling the inspection of the connectivity at the physical level, such as RSRQ (Reference Signal Received Quality) and RSRP (Reference Signal Received Power). For different providers, we compare the achieved speeds from measurements that indicate the same LTE connectivity quality and expect that in case no throttling is in place, at the same RSRP/RSRQ value we get approximately the same value of the achieved speed for each of the providers. Figure 1 shows that this is not the case, and that there is an observable difference among the ISPs—"Provider 2" exhibiting the best performance.

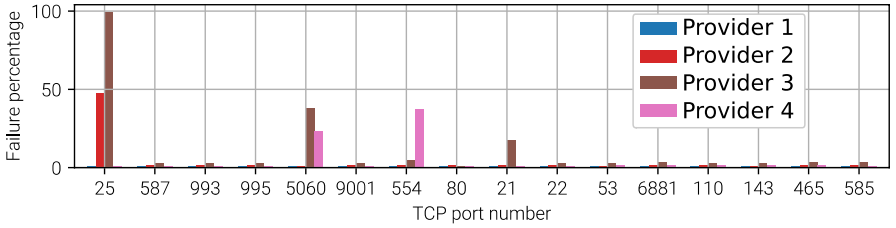

Fig. 2 TCP port blocking results for four major Slovenian ISPs. High failure percentages indicate that providers 2 and 3 often block port 25, Providers 3 and 4 port 5060, while Provider 4 blocks port 554 approx. 40% of the time

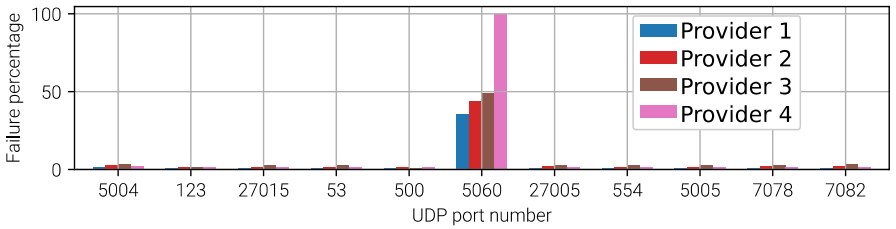

Fig. 3 UDP port blocking results for four major Slovenian ISPs. All ISPs occasionally block port 5060

The comparison, however, is not a strong indicator of Net neutrality violations, since throttling is legitimate, as long as it is in line with the rates that the subscribers pay for. Without information on the actual user-operator contracts, we have no means of discriminating between the legitimate and non-legitimate throttling. Furthermore, our approach captures a single dimension of the context—connectivity quality. Other dimensions can skew the ISP comparison results. For instance, users of Provider 2 might be bound to contracts that include the latest mobile devices, which further contribute to higher speed measurements. While we have information on device types and have implemented a method for enrichment analysis (see Sect. 6), the sparsity of the data prevents deeper multidimensional analysis in this case. Finally, cofounding parameters, not captured by the measurement system, may impact results shown in the figure. Such a parameter could be the bandwidth of the channel used by different ISPs. Wider channels would, for the same connection quality indicators, enable higher download and upload speeds. This information is not obtainable from purely crowdsourced measurements. Even measurements with specialized testing equipment employed by regulators may fail to capture all the subtleties of the connection. For example, the bandwidth measured by such equipment would not capture the restrictions on the bandwidth use that Mobile Virtual Network Operators (MVNOs) have when using the host network infrastructure.

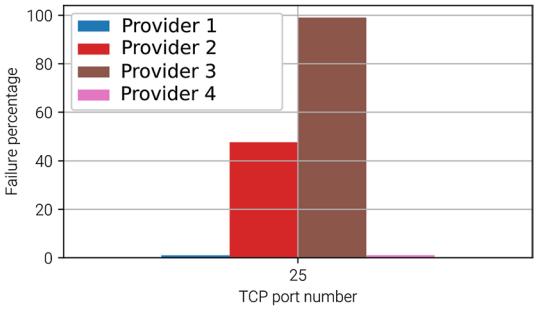

Fig. 4 TCP port 25 blocking by major Slovenian operators

4.2 Port Availability Analysis

Together with each speed test, AKOS Test Net mobile Android application executes a set of tests where it attempts to transfer data to a remote server under the regulator's control, using TCP or UDP flows. The transfer is attempted over 16 TCP and 11 UDP ports, corresponding to common network applications (e.g. SMTP, SSH, FTP, etc.). The result of each sub-test is a binary—port open or closed—information. We aggregate the outcomes of all tests over different port numbers and for different providers and show them in Fig. 2 for TCP and Fig. 3 for UDP flows.

The figures show, on the average, a very low failure rate, indicating that for the majority of ports, no port blocking is in place. However, we see that tests on certain ports, such as TCP ports 25, 554, and 5060, and UDP port 5060 often experience failures. Port 25 is used for SMTP, i.e. email posting, port 554 for video content, while port 5060 is often used for Voice-over-IP connections. The difference between the success rates of tests on these ports and on other ports in the database is statistically significant. Since the same NRA-authored "service" listens at each of the ports, we take the above result as a strong indicator that the ports are indeed being blocked.

We now zoom into examine port blocking practices among different ISPs. In Fig. 4 we show the TCP port 25 test success rate over operators for which we have at least 500 port availability measurements per operator. There is a statistically significant difference between the test success rate of different operators. Clearly, subscribers of Provider 3 experience port 25 blocking, almost without an exception. While this often does not happen in practice, there is a slight chance that the blocking is performed by another ISP, whose infrastructure lies on the path from Provider 3 to the test server. Intermediate blocking is more common in local networks. For instance, our analysis of measurements taken while the devices were connected to WLAN indicates that a national academic network blocks ports more frequently than any other ISP—being familiar with our department's IT service rules, we were not surprised by this finding. Finally, we should note that ISPs often block port 25 in order to force users to revert to a more secure SMTPS protocol, usually hosted at port 465.

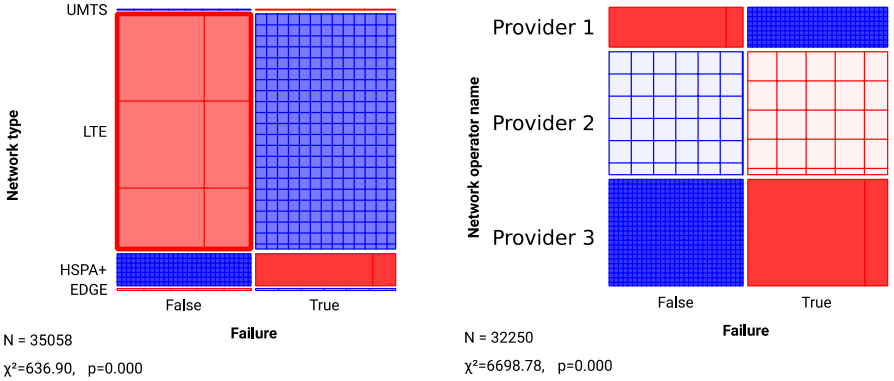

Fig. 5 Sieve diagram showing the prevalence of different network connectivity types and different ISPs in unsuccessful VoIP measurements

Table 2 Enrichment analysis for VoIP tests in AKOS Test Net data. Certain combinations of parameter values, such as Provider 3 and HSPA+, tend to be over-represented in successful measurements

Network operator name	Network type	Count	Count-class	Enrichment	p-value
Provider 3	HSPA+	1772	1453	1.675	4.93e−193
Provider 3	LTE	10,817	7144	1.349	0.00
Provider 2	HSPA+	2018	1538	1.557	2.27e−147
Provider 2	LTE	12,264	5498	0.916	3.18e−11
Provider 1	HSPA+	744	1	0.003	4.32e−11
Provider 1	LTE	4309	1	0.0005	3.54e−11

4.3 VoIP Test Performance

To assess the performance of VoIP, together with speed and port availability tests, AKOS Test Net mobile app conducts measurements with VoIP-like artificially generated traffic. The test then examines the inbound/outbound packet delivery rate and packet inter-arrival time jitter. In case of non-zero packet delivery rate and the average jitter lower than 50 ms, the test is marked as successful.

The test success, however, can depend on a number of parameters. For instance, received signal strength, network type, or even the device's capabilities can impact the result. Orange data mining suite allows us to analyze the data along different dimensions. In Fig. 5 we show the so-called "sieve diagram" of VoIP test results. The diagram depicts the actual and the expected frequencies of different parameter values in the dataset. As parameters we use the connection type (UMTS, LTE, HSPA+, EDGE) in the left figure and the ISP (Provider 1, Provider 2, and Provider 3) in the right figure. Blue color in the figure indicates that the given parameter value is over-, while red color indicates that the value is under-represented in the data, compared to the expected, uniform, distribution

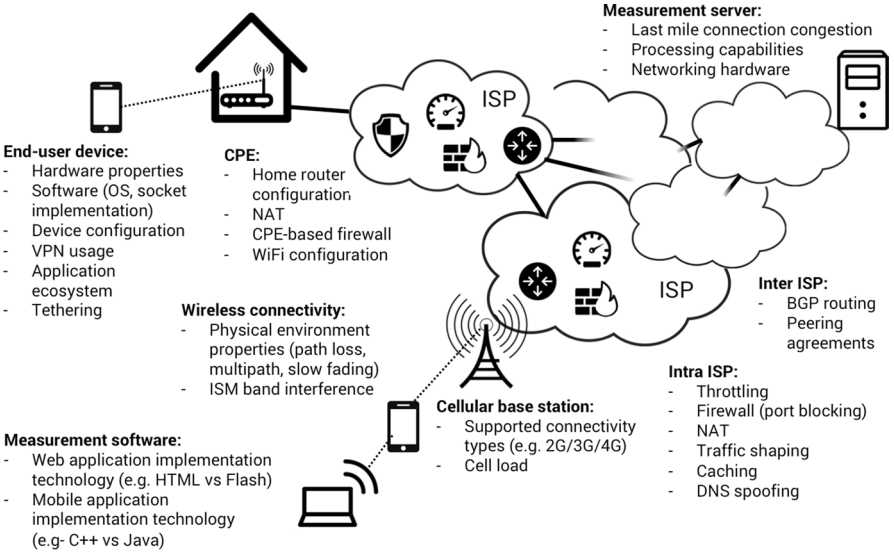

Fig. 6 Overview of factors that may impact network measurement results

(i.e. failures are not a priori considered prevalent in a certain network type or with a certain ISP). The density of the network for the given value-result combination indicates the intensity of over/under representation. From the figures it follows that HSPA+ connection type and Provider 3 tend to be over-represented in successful VoIP tests (Failure = "False"). However, there is still a question on how different combinations of the parameters impact the results.

Hypergeometric test is a statistical method that allows us to quantify how much different parameter value combinations enrich an outcome from a dataset. We extend Orange data mining suite with *Significant Groups* widget that, among other functionalities, contains the hypergeometric test. In Table 2 we show the results of the analysis. *Enrichment* column indicates the over-representation level of the given parameter combination for *VoIP test failure* = *False* data points. We observe that the combinations of Provider 3 and HSPA+ or LTE, and Provider 2 and HSPA+ tend to be over-represented in successful measurements. The p-values (all lower than 0.001) indicate that the differences are statistically significant.

In the context of Net neutrality violation detection, the hypergeometric test allows us to single out the role of different confounding variables and precisely pinpoint the impact of an ISP on the results.

5 Discussion

The analysis we conducted in the previous sections points to serious limitations of Net neutrality violation detection that is based solely on crowdsourced mobile broadband measurements conducted with a variety of end-user equipment. Due to the lack of other

sources of information, in particular from the ISP side, we could not attribute measured differences among speeds that end-users observe at different ISPs to illegal throttling, nor could we ascertain that port blocking is indeed performed by the last-mile ISP.

To set guidelines for the implementation of a reliable Net neutrality violation detection system, we now look into a holistic context of Internet connectivity. In Fig. 6 we depict a full high-level schema of Internet connectivity showing how end-devices, such as mobile phones and laptops, connect through last-mile connections to ISPs, which are themselves interconnected. In the figure we also show the measurement infrastructure, which includes measurement software (often, just like in the case of AKOS Test Net, installed on end-user devices) and measurement servers. Different entities lie on the path from a consumer to a measurement server. For instance, a mobile measurement app's speed test results may be impacted by a whole range of factors: the operating system of the user device (e.g. how the sockets are actually implemented, what kind of the TCP congestion control protocol is implemented, etc.), possible VPN tunneling that the user has in place, contention with other applications running in parallel on the device, the type and the properties of the wireless link, wireless interference, congestion at the base station, ISP policies, such as throttling, port blocking, NAT-ing, traffic shaping, content caching, etc., interconnection of the ISPs, with factors such as BGP routing policies, peering agreements, etc., and finally, server capabilities and the server-to-Internet connection properties.

Consequently, a system for Net neutrality violation detection should:

- Encompass a wide range of measurements enabling a holistic view of network performance;
- Include statistical methods that can isolate the role of individual parameters on the measurement results.

AKOS Test Net data we analyzed contains a range of measurements related to connection speed and QoS. However, as shown in Sect. 5, the measurement system fails to pinpoint the culprits of Net neutrality violations. Moreover, AKOS Test Net data are obtained without the supervision of the measurement points and control over the measurement times and locations. Thus, the data contains uneven distribution with respect to the connectivity type, operator, and other parameters. Furthermore, the measurement system cannot ensure that the operators are not recognizing and reacting to measurement campaigns. With a holistic approach to measurements, that includes a wider range of measurements, tighter control over the measurement equipment, fusion of various sources of data, their in-depth statistical analysis, and, finally, adaptive, targeted and iterative measurements, we can improve the reliability of the measurement system for Net neutrality violation detection.

5.1 Comparison of NRAs' Approaches to Net Neutrality Violation Detection

Various national regulators, often with the assistance of academic groups and industry players, have designed and deployed systems for Net neutrality violation detection. In Table 3 we juxtapose selected documented approaches with the approach

Table 3 Comparison of different national regulators' approaches to Net neutrality monitoring

Approach	Measurement and analysis	Abilities	Note
Bredbeitung (Germany)	Crowdsourced measurements performed on demand with customers' own devices. Analysis performed on demand after users' complaints	TH-Y PB-N SD-Y TA-N CV-Y	To ensure robust results, the approach requires at least 20 measurements in 2 days, conducted over the same reliable fixed connection. Mobile crowdsourced measurements are not considered robust enough for further Net neutrality violation analysis TH and CV: the assessment is based on the comparison between the plan details reported by the customers (which might be prone to errors) and the measured statistics; additionally, no ISP-side information about temporary throttling is available SD: inference is not automated but handled on an individual case basis after sufficient complaints
Adkintun (Chile)	Custom firmware routers for crowdsourced fixed connection measurements and an app for mobile broadband measurements The analysis methodology defined by the University of Chile's researchers defines the requirements for a statistically representative sample	TH-Y PB-N SD-N TA-N CV-Y	Active measurements of throughput, delay, and packet loss rate are periodically performed between end-hosts and several measurement hosts distributed across the country During the statistical analysis, bootstrap resampling is harnessed to identify confidence intervals for observations made from the measurements TH and CV: the assessment is based on the comparison between the plan details reported by the customers (which might be prone to errors) and the measured statistics; additionally, no ISP-side information about temporary throttling is available
Ofcom (UK)	ISPs provide sync speeds of each active line For fixed connections custom hardware-based measurements. For mobile connections: single device tests while driving and walking, with a mixture of indoor and outdoor locations Experience feedback collected via user panels Yearly Connected Nations report presents descriptive statistics. No further statistical analysis methodology defined	TH-Y PB-N SD-Y TA-N CV-Y	SD: For fixed connections Netflix streaming performance is measured (in addition to average daily disconnections, Web browsing speed, latency, packet loss, DNS resolution, DNS failure, and jitter). Other services are not mentioned in the report TH and CV: while ISP-side information potentially enables automated detection of throttling that violates contractual agreements, this is not utilized by the proposed approach

Table 3 (continued)

Approach	Measurement and analysis	Abilities	Note
Measuring Broadband America (USA)	QoS characteristics (e.g. download/upload speed, latency, loss) are measured. Fixed broadband measurements are made using hardware probes, whereas mobile broadband measurements are made via an app installed by a panel of volunteers Aggregate results presented in a yearly report. Further statistical analysis methodology not defined	TH-Y PB-N SD-Y/N TA-N CV-Y	The main goal of the approach is to ensure transparency and the compliance of ISPs with the advertised performance metrics SD: The test periodically measures the total time to request and receive web pages from nine popular websites. Differentiation among other services is not detectable
AKOS Test Net (Slovenia)	Crowdsourced measurements via a mobile app or via a Web page Statistical analysis methodology not defined	TH-Y PB-Y SD-Y/N TA-Y/N CV-N	SD: VoIP test, results of which are analyzed in the previous section, reveals the difference in VoIP transfer performance over different ISPs. Differentiation of other kinds of traffic and differentiation of different services over the same ISP are not detectable TA: AKOS test net contains a test that sends a malformed message to the test server's port 25, to which the server does not react. If a response is still detected, that indicates a non-transparent proxy. A similar test is conducted with a JPEG-encoded image—signatures of the original and the received image are compared to detect potential tampering. Other aspects of TA, such as transcoding of media in general and TCP splitting, are not detectable

"Y"/"N" indicate the whether the approach has the ability ("Y") or inability ("N") to detect different aspects of Net neutrality violations: throttling (TH), port blocking (PB), service differentiation (SD), traffic tampering (TA), and contract violations (CV)

taken by the Slovenian national regulator. In particular, we are interested in the systems' ability to assess the KPIs we defined in Sect. 4, as well as the (in)ability to detect contract violations and tampering with the network traffic. The assessment in the table depicts the best case, i.e. the ability of an approach to detect throttling (TH), port blocking (PB), service differentiation (SD), traffic tampering (TA), and contract violations (CV), if a sufficient number of reliable measurements is available.

The above table demonstrates common limitations of explanatory power of the existing approaches to Net neutrality violation detection. Most notably, detecting differentiation across a range of services and detecting traffic tampering (e.g. media transcoding) remain outside of the scope of the currently employed approaches. Furthermore, the ISP-side information about the actual contractual agreements and temporary throttling are not collected. Finally, statistical methodology is almost never defined and the potential of advance data mining methods remains unharnessed. In Sects. 5.2, 5.3, and 5.4 we address these issues, before proposing a full-fledged holistic Net neutrality violation detection system in Sect. 6.

5.2 Case for Additional Crowdsourced Measurements

Network measurements for Net neutrality violation detection, on the long term, have to follow changes in the way the operators treat the traffic. On the short term, however, the data has to uncover all the latest possible means of traffic (mis)handling by an ISP. The measurement system we analyzed should be extended with:

- **Throttling detection for different applications and content types:** based on URL patterns or deep packet inspection (DPI) the operators may selectively throttle certain data flows. Methods for discovering such discrimination among flows are still at the research stage (see Sect. 2). The most common approaches rely on the relative discrimination detection, where a network traffic flow is recorded and replayed through different ISPs [20], with or without a VPN [27], or with shuffled content [24]. A statistical analysis of the measured performance parameters could then be used for violations detection. An alternative approach relies on passive observation and comparison of similar flows in networks of different operators, or the comparison of similar flows of different applications/parameters [25]. Yet, this approach has been demonstrated in a relatively uniform wired-connection setup of the PlanetLab research network and would unlikely be able to cope with a wide range of factors, such as end-user device types, mobility, location, rare applications, and non-uniform ISP popularity observed in real-world MBB deployments. Finally, a recently published paper by Raida et al. describes a level shift detection method for throttling detection [34] which shows promising performance in case fine-grain measurements are available.
- **Transcoding detection:** measurements conducted by Kakhi et al. show that certain ISPs (in the specific example Boost Mobile and Sprint) occasionally replace high-quality images and videos with their lower-quality (transcoded) counterparts [27]. Transcoding can be detected if the signature (e.g. a content hash) of

the original media file is compared with the received media signature. AKOS Test Net already contains a test that fetches a JPEG image from the test server and compares its signature with the expected original signature. However, a comprehensive test should include a number of images of different sizes and formats, as well as a video content. Furthermore, tests should be ran from different sources, in order to prevent ISPs from detecting measurement campaigns.

- **TCP splitting detection:** while not necessarily limiting, TCP splitting is a method that interferes with the expected flow behavior and can be applied discriminatively, thus is a borderline Net neutrality violation practice. Splitting can be detected by the comparison of the TCP handshake, i.e. arrival times of SYN, ACK, and SYN-ACK packets, at the two ends of the connection, as well as through the identification of middlebox *fingerprints* in TCP/IP headers of the transmitted packets [28].
- **Networking equipment misalignment:** by itself, misalignment of the equipment is not violating Net neutrality. However, in certain examples, for instance if the parameters of a duplex connection are not properly agreed upon by network interface cards, it can lead to poor data transfer performance that can be misinterpreted as throttling. NDT tests[8] implemented within AKOS Test Net already include misalignment detection, yet, due to optional inclusion of such a test very few test instances were indeed conducted during the two-year period we have analyzed.
- **Advanced transfer limitation detection:** here we group Net neutrality violation techniques that are rather challenging to detect, as they often require flow state monitoring, yet are observed in practice. Examples include:
  - Sending TCP RST packets, which leads to connection termination, when BitTorrent flow exceeds certain data flow size limit [35];
  - Blocking HTTP requests for Flash video content larger than 20 MB [36];
  - Capturing and forwarding search queries towards ISP-managed servers;
  - Routing the differentiated traffic through slower links.

Detailed flow parameter monitoring and interactive iterative analysis represent the first step towards advanced Net neutrality violation techniques detection.

5.3 Case for Supervised Measurements

Factors outside the NRA's control, such as the end-devices' properties, applications running in parallel to the measurement app, users' locations, and other factors, impact crowdsourced measurements. Obtaining more reliable measurements results is possible with the use of calibrated dedicated measurement equipment. Such an approach is already employed by FCC in the USA and Ofcom in the UK, which use specialized broadband quality and speed measurement equipment developed by SamKnows. In these particular cases, the equipment is installed at

[8] https://www.measurementlab.net/tests/ndt/.

end-users' homes and is used for static broadband measurement only. A drawback of the approach its scalability—the cost of equipment limits the installation to a fraction of possible users/locations. Consequently, a combined approach including large-scale crowdsourced and well-controlled dedicated installations is discussed in the next section.

5.4 Case for Transparency, Reporting, Integration with ISPs' Information Systems, and User Feedback Monitoring

Independent measurement campaigns orchestrated by NRAs cannot, due to a large number of entities on the path from an end-user towards a measurement server (Fig. 6), provide a reliable estimate of all Net neutrality violation issues. Instead, it is necessary for NRAs to work together with ISPs, in particular by:

- **Requiring transparency and reporting** from the ISPs side. Certain countries already require from ISPs to provide information about offered services and the employed means of traffic management. For instance, in Chile the regulatory framework requires the ISPs to provide quarterly reports on:

  - The advertised upload/download speeds, data transfer limits, and the quality of the delivered service;
  - Infrastructure contention ratio;
  - Technical indicators;
  - Replacement time of the service;
  - Quality and availability of the link;
  - Traffic- and network-management practices, including their characteristics and effects on the service provided to users. The information includes the types of applications, services and protocols that are affected;

  Besides the above, our analysis points out to *peering agreements*, as an additional valuable piece of information that should be provided by ISPs. We note that peering agreements may represent important strategic information for ISPs exposure of which might impact one ISPs competitive advantage over another ISP. Consequently, NRAs should ensure that the information is revealed exclusively to the regulator and for the purpose of Net neutrality violation detection.

- **Integrating ISPs' information systems with the measurement framework** Our analysis also points to the lack of user-level information crucial for Net neutrality violation inference:

  - Subscription package of the connection under consideration (e.g. nominal download/upload speed);
  - Subscription account state (quota size and depletion);
  - Throttling (temporary throttling, throttling of different apps/protocols, etc.);

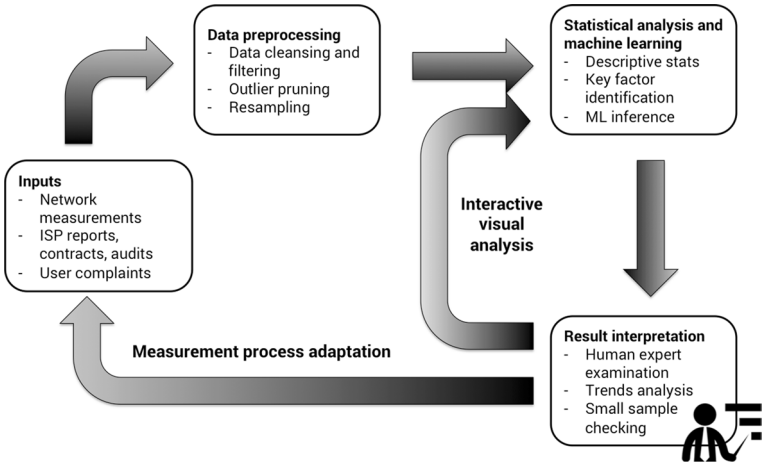

Fig. 7 Iterative approach to Net neutrality violation detection

This information can be provided solely by the ISP. An application programming interface (API) at the ISPs' side, accessible by the NRA's measurement tools, and providing limited information on the subscribers' accounts, would enable seamless integration of the above information and the independent measurements.

- **User feedback monitoring** Independent measurements enable Quality of Service (QoS) monitoring. A discrepancy between the user-observed Quality of Experience (QoE) and QoS is possible, especially if not all relevant connection parameters are included in the measurements. Enabling users to report issues with the experienced connectivity would help NRAs guide further measurements and single out Net neutrality violation cases that impact users the most. Reporting systems are already in place, installed either by national NRAs, e.g. in Germany,[9] or by third parties, e.g. Respect My Net initiative.[10]

6 Iterative Machine Learning-Based Approach to Net Neutrality Violation Detection

The overview of both related academic work, as well as worldwide NRA practices, together with our analysis of crowdsourced measurements collected by the Slovenian NRA, provide a basis for an approach to Net neutrality violation detection presented in this section. The approach is grounded in the statistical analysis and guided iterative analysis with a human expert in the loop. The main phases of the data processing are shown in Fig. 7:

[9] https://breitbandmessung.de/.
[10] https://respectmynet.eu/.

- *Input data consist of results acquired through various measurements*, e.g. crowdsourced measurements, such as the ones collected by AKOS Test Net, fine-grain wireless protocol measurements collected via drive tests,[11] and measurements with dedicated equipment. Besides the target variables, such as RTT, network speed indicators, port reachability, and similar, the measurements should contain as much contextual information that may impact the measurement results—network connectivity parameters (e.g. type, signal strength, frequency, etc.), measurement device information (device model and make, battery information, CPU usage, etc.), location information, and others—as possible. Extra care should be taken to ensure mergeability of data coming from different sources. Finally, input data should include ISP-side information about the subscriptions, throttling, and peering agreements.
- *Data preprocessing* is a necessary step in mobile broadband measurement analysis. Sudden service disruptions due to bursty network usage, routing anomalies, and other effects, can impact aggregate values and lead to incorrect reasoning about network phenomena. At minimum, outlier measurements should be filtered out, while empirical distribution sampling may be appropriate when imbalanced data prevent further statistical analysis.
- *Statistical data analysis and machine learning* includes descriptive analysis with which easy-to-spot patterns can be detected, but also more sophisticated machine learning methods. The former includes, for instance, standard deviation, mean, maximum, and minimum calculation and comparison among these parameters computed for different groups of measurements (e.g. measurements taken at different nodes, at the same node at different times of day, at the same node with different providers, etc.), using parametric or non-parametric tests (e.g. t-test, Mann–Whitney U test, etc.) we can quantify the difference and our confidence that the observed differences are indeed significant. Enrichment analysis (see Sect. 6.1) can then be used to then identify factors that cause differences in measured values among groups of measurements. More advanced models can be built to predict measurement values and detect deviations from the expected behavior. Regression models that take the context of the measurement, including the ISP-related information, location, signal strength, etc., are appropriate for predicting numerical measurement results (e.g. RTT, download/upload speed), while classification models should be built for predicting categorical observations (e.g. port open/closed). The temporal nature of the data calls for quick detection of anomalies, and for that concept drift detection methods, such as ADWIN can be applied [37]. While deviations from the predicted classification/regression, or detected concept drifts provided by the above models can be used for semi-automatic Net neutrality violation detection, the final verdict should be provided by a human expert who has a thorough understanding of all confounding factors that might have impacted the results. Certain aspects of statistical analysis among measured

[11] An example of drive test software is Rohde & Schwartz ROMES4 https://www.livingston-products.com/products/pdf/156521_1_en.pdf.

Table 4 Net neutrality violation detection problems and corresponding techniques that tackle them

Problem	Technique	Solution details
Unknown baseline performance	Data-driven machine learning models	Crowdsourced measurements are used to devise models that relate the measurement context (e.g. ISP, device model, connection type, signal strength, etc.) with the measured performance. These models are constructed from the best performing measurements in each contextual group and provide the baseline against which further measurements are assessed—large deviations indicate potential traffic shaping/throttling
Discerning between legitimate and unauthorized throttling	ISP-side information on individual connection properties	ISPs provide APIs that NRAs can query and obtain information about the actual subscriber line speed, subscription quotas, and other relevant information. The same API is used to obtain information about temporary throttling due to surges, flash crowds, and similar events
Middleboxes are not detected	State-of-the-art academic solutions for middlebox detection employed	In collaboration with academic researchers, state-of-the-art approaches to middlebox detection, including transcoding and TCP splitting, are installed; the measurements are conducted in sufficient amounts to arrive to statistically valid conclusions about the middlebox existence
Identifying factors leading to measurement deviations	Statistical comparison between well-performing and substandard measurements	Statistical parametric or non-parametric tests (e.g. t-test, Mann–Whitney U test, etc.) conducted between different groups of measurements, with enrichment analysis used to identify the discerning factors (see Box below)
Crowdsourced measurements are unreliable, custom hardware measurements do not scale	Hierarchical crowdsourced—custom hardware measurements	Crowdsourced measurements, together with the users' feedback, represent the basis for a high-level warning system; once a warning is triggered, targeted detailed measurements with custom hardware are conducted in selected area(s)

Table 4 (continued)

Problem	Technique	Solution details
Net neutrality violations may be temporal and sophisticated	Automated anomaly detection as an alarm	Automated detection using ADWIN, Page-Hinckley and other anomaly detection methods are used to warn a human expert about the need for further inspection
Net neutrality violation methods are evolving	Academic research and expert in the loop	Selected techniques from recent research efforts in Net neutrality are integrated in the holistic system. Networking experts are kept in the loop via an interactive dashboard that enables small-scale studies and trend visualization

flows has already employed in measurement analysis systems such as POPI [22], NANO [25], DiffProbe [21], and Packsen [23].
- *Result interpretation with a human expert in the loop* is a necessary step after the analysis. The complexity of the influence of different infrastructure, software, user-behavior, ISP policies, and other factors (see Fig. 6) call for a careful interactive visual inspection of the results of the statistical analysis.

The proposed approach is tailored to tackle some of the key open issues in the area Net neutrality violation detection. In Table 4 we list some specific problems and the corresponding techniques that address these problems.

6.1 Enrichment Analysis

Enrichment analysis—"which (combinations of) attribute values lead to the measured outcome" remains one of the most important questions in mobile broadband data analysis. For example, we might discover that connection type, phone model, the operator, or a combination of the above leads to unsuccessful VoIP tests. Hypergeometric test is a statistical method appropriate for such an analysis. We have implemented an Orange data mining suite module called "Significant Groups" containing the hypergeometric test. The module contains tools, such a t-test, Flinger-Killeen test, Mann–Whitney U test, and Gumbel distribution test, for the identification of significant factors in the case of categorical or numeric outcome variables, allowing us to answer questions such as "How do phone models and/or connectivity types impact the measured download speed?". The implementation is available as a part of our RICERCANDO open source framework.

6.2 Evolution of the System

Iterative analysis-based Net neutrality violation detection system sketch in the previous section requires constant updating, as measurement and mining techniques evolve. In addition, we should be aware of the following factors that may impact the system reliability:

- *Changes in measurement context.* Connection links and their quality, peering agreements among ISPs, network congestion rates, protocol types and their implementations, network and end-user devices' hardware, the number and the behavior of the users, and a range of other factors may change, either temporarily or permanently, either periodically or as a trend, and affect the recorded results. Furthermore, the phenomena Net neutrality violation detection systems try to capture are themselves dependent on the context. For instance, Goel et al. show how certain ISPs use TCP connection splitting proxies, yet once these proxy servers become overburdened the splitting is not employed any more [28].
- *Measurement self-interference.* Active probing can affect the measurement results. As an example, frequent measurement via AKOS Test Net mobile app

may deplete a user's data plan, and thus, result in download/upload throttling from the ISP's side.

- *ISPs' adaptation to measurement systems.* With further pressures to operate in accordance with Net neutrality regulations, ISPs will likely adapt and ensure that NRAs' measurements do not reveal any violations. In discussion with another European institution working on Net neutrality violation detection through crowdsourced measurements we were presented with findings that demonstrate an operator's ability to detect an NRA's measurement campaign traffic and open an "unlimited rate" new Access Point Name (APN) for such traffic. Furthermore, crowdsourced measurements enable efficient collection of a large number of measurements, but without much control over the source of the measurements. Thus, ISPs could, in theory, include their own devices with prioritized subscription accounts in the existing NRA's measurement campaigns.
- *Detecting a moving target.* Net neutrality violation detection systems are still in their nascency. Known examples of violations are at the moment rather simple—TCP/UDP port blocking, URL blacklisting, zero rating, and similar. A development of detection tools will likely result in an arms race between NRAs and ISPs. As an example, consider VoIP (SIP) protocol blocking. Instead of TCP/UDP port 5060 blocking, the protocol can be blocked by more advanced techniques, such as packet size and packet interarrival time monitoring and classification (into "VoIP" or "other traffic"). AKOS test Net app, for example, already contains a port blocking test, yet, in the current implementation the app would not be able to detect SIP blocking founded on more advanced techniques listed above. *Analogous to cybersecurity, we believe that Net neutrality violation detection requires a constant evolution of measurement and detection techniques*.

6.3 Case Study—Throttling Detection

We now discuss a hypothetical scenario inspired by the inconclusive analysis we presented in Sect. 4.1. We are again interested in detecting download throttling, yet, we show how the approach sketched in this section successfully tackles the limitations of the crowdsourced measurements we initially relied upon.

We assume an approach based on *machine learning modeling of the expected download rates*, obtained via crowdsourced measurements, when the measurements are taken in different contexts. A context is defined, above all, by the physical parameters of the connection, e.g. signal strength, signal quality indicators, frequencies used, bandwidth, but also measurement hardware, e.g. a phone's model and make, and software, e.g. operating system, CPU usage, etc. Data collected over a longer period of time and sufficiently stratified over the contextual parameters, are filtered for outliers and then used to build a model, for instance, the one based on a quantile regression forest [38], that predicts the expected download speed in a given context. The quantile regression forest model is suitable for this purpose as it predicts the dependent variable value even if the context has not been observed before and can take into account the top, e.g. 10^{th} percentile of measurements, thus, produces a reliable model even if the training data contain occasional throttling instances.

With the model in place, further crowdsourced measurements provide a bird's eye view of the performance. As before, these measurements should contain download speed measurements complemented with the context measurements. Furthermore, through APIs provided by ISPs, each measurement should be paired with the information on a user's package details (e.g. max/min speed defined in the contract, whether the connection was throttled due to data caps being exceeded, etc.). Next, the actual download speed measurements are compared to model-based predictions that take the context as the input. If statistical tests (e.g. t-test) show significant deviations from the expected results, and if those deviations cannot be explained by the limitations of the subscription package used, we proceed with the identification of the causes for the discrepancy. We perform a hypergeometric test to single out factors that are over- or under- represented in the anomalous data. Such factors may include ISPs, geographical regions, and others. Suppose that combinations of ISPs and regions are identified.

Measurements with dedicated measurement equipment are not scalable yet are more reliable than crowdsourced measurements conduced on users' end devices. Thus, with the potential culprits identified, we proceed with the deployment of dedicated measurement equipment in target regions and for target ISPs. Additional measurements conducted with such equipment can now include traceroute-like download measurements to identify hops at which throttling happens [20], tests with different traffic types, packet size distributions, and other factors. At this point communication with ISPs should be intensified to obtain additional info that may be necessary for pinpointing the reason for underperforming downloads, such as the information on peering agreements. An expert in the loop is in charge of guiding these small-scale measurements and interpreting the results until the true root cause of throttling is identified.

7 Conclusions

In this paper we have presented the analysis of more than 2 years of broadband measurement data collected by the Slovenian national regulator, with the goal of inferring the limits of crowdsourced approaches for Net neutrality violation detection. Our analysis, one of very few conducted on country-wide data, demonstrates low inference power when the crowdsourced approach is used in isolation. Therefore, in this paper we also discuss augmentations, such as measurements with dedicated calibrated equipment, additional measurements of advanced discrimination mechanisms that rely on middleboxes, such as transcoding, integration with ISPs' information systems in order to obtain up-to-date information on subscriber accounts, and the inclusion of a user feedback monitoring system. In the paper we also emphasize the role of a statistical analysis, in particular when it comes to the identification of factors that may contribute to the observed violations. Finally, we note that a successful holistic analysis requires an iterative approach, where initial measurements serve as a warning signal guiding an expert's focus, while successive targeted probing might be needed to pinpoint the exact extent and the reason for the observed anomaly.

Acknowledgements The author would like to thank Ivan Majhen and Miha Janež for their help with the data analysis, Janez Sterle for discussions about AKOS Test Net system and result interpretation, Narseo Vallina-Rodriguez for discussions about Net neutrality violation detection, and Urban Kunc for his help with the interaction with the measurement system. The work was partly funded by the Agency for Communication Networks and Services of the Republic of Slovenia (AKOS) and by the European Union's Horizon 2020 research and innovation programme underGrant Agreement No. 644399 (MONROE) through the open call project RICERCANDO. The views expressed are solely those of the author.

References

1. Waverman, L., Meschi, M., Fuss, M.: The impact of telecoms on economic growth in developing countries. Vodafone policy Pap. Ser. **2**(03), 10–24 (2005)
2. Manyika, J., Roxburgh, C.: The great transformer: The impact of the Internet on economic growth and prosperity. McKinsey Glob. Inst. **1**, (2011)
3. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Futur. Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
4. Hart, J.A.: The net neutrality debate in the United States. J. Inf. Technol. Polit. **8**(4), 418–443 (2011)
5. Ganley, P., Allgrove, B.: Net neutrality: a user's guide. Comput. Law Secur. Rev. **22**(6), 454–463 (2006)
6. Allen, J., Daly, J. A., Marcus, S., de Antonio Monte, D., Woolfson, R.: Study on net-neutrality regulation. (2017)
7. Leiner, B.M., et al.: A brief history of the Internet. ACM SIGCOMM Comput. Commun. Rev. **39**(5), 22–31 (2009)
8. Crowcroft, J.: Net neutrality: the technical side of the debate: a white paper. ACM SIGCOMM Comput. Commun. Rev. **37**(1), 49–56 (2007)
9. Wu, T.: Network neutrality, broadband discrimination. J. Telecomm. High Tech. L. **2**, 141 (2003)
10. European Commission: Regulation (EU) 2015/2120 of the European Parliament and of the Council. Off. J. Eur. Union **L 310/1**, (2015)
11. BEREC: BEREC Guidelines on the Implementation by National Regulators of European Net Neutrality Rules. (2016)
12. Weber, M., Svedek, Jukic, V. Z., Golub, I., Zuljevic, T.: Can HAKOMetar be used to increase transparency in the context of network neutrality? In: 2013 12th International Conference on Telecommunications (ConTEL), pp. 309–316. (2013)
13. Ofcom: UK broadband speeds—the performance of fixed-line broadband delivered to UK residential consumers. (2010)
14. Miorandi, D., Carreras, I., Gregori, E., Graham, I., Stewart, J.: Measuring net neutrality in mobile Internet: towards a crowdsensing-based citizen observatory. In: 2013 IEEE International Conference on Communications Workshops (ICC), pp. 199–203. (2013)
15. Bustos-Jiménez, J., Fuenzalida, C.: All packets are equal, but some are more equal than others. In: Proceedings of the Latin America Networking Conference on LANC 2014, p. 5. (2014)
16. Garrett, T., Setenareski, L.E., Peres, L.M., Bona, L.C.E., Duarte, E.P.: Monitoring network neutrality: a survey on traffic differentiation detection. IEEE Commun. Surv, Tutorials (2018)
17. Dischinger, M., Marcon, M., Guha, S., Gummadi, P. K., Mahajan, R., Saroiu, S.: Glasnost: enabling end users to detect traffic differentiation. In: NSDI, pp. 405–418 (2010)
18. Gregori, E., Luconi, V., Vecchio, A.: NeutMon: studying neutrality in european mobile networks. In: IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). pp. 523–528. (2018)
19. Gregori, E., Luconi, V., Vecchio, A.: Studying forwarding differences in european mobile broadband with a net neutrality perspective. In: European Wireless 2018; 24th European Wireless Conference, pp. 1–7. (2018)
20. Zhang, Y., Mao, Z. M., Zhang, M.: Detecting traffic differentiation in backbone ISPs with NetPolice. In: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement, pp. 103–115. (2009)
21. Kanuparthy, P., Dovrolis, C.: Diffprobe: detecting ISP service discrimination. In: 2010 Proceedings IEEE INFOCOM, , pp. 1–9. (2010)

22. Lu, G., Chen, Y., Birrer, S., Bustamante, F.E., Li, X.: POPI: a user-level tool for inferring router packet forwarding priority. IEEE/ACM Trans. Netw. **18**(1), 1–14 (2010)
23. Weinsberg, U., Soule, A., Massoulie, L.: Inferring traffic shaping and policy parameters using end host measurements. In: 2011 Proceedings IEEE INFOCOM, pp. 151–155. (2011)
24. Ravaioli, R., Barakat, C., Urvoy-Keller, G.: Chkdiff: checking traffic differentiation at internet access. In: Proceedings of the 2012 ACM Conference on CoNEXT Student Workshop, pp. 57–58. (2012).
25. Bin Tariq, M., Motiwala, M., Feamster, N., Ammar, M.: Detecting network neutrality violations with causal inference. In: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, pp. 289–300. (2009)
26. Choffnes, D., Gill, P., Mislove, A.: An empirical evaluation of deployed dpi middleboxes and their implications for policymakers. In: Proc. of TPRC, (2017)
27. Molavi Kakhki, A., et al.: Identifying traffic differentiation in mobile networks. In: Proceedings of the 2015 Internet Measurement Conference, pp. 239–251. (2015)
28. Goel, U., Steiner, M., Wittie, M. P., Flack, M., Ludin, S.: Detecting cellular middleboxes using passive measurement techniques. In: International Conference on Passive and Active Network Measurement, pp. 95–107. (2016)
29. Schaurich, V. G., de Carvalho, M. B., Granville, L. Z.: ISPANN: A policy-based ISP auditor for network neutrality violation detection. In: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), pp. 1081–1088. (2018)
30. Specure: SPECURE NetTest—overview of the technical setup, data collection methodology and reporting capabilities. (2015)
31. Pejovic, V., Majhen, I., Janez, M., Zupan, B.: RICERCANDO: data mining toolkit for mobile broadband measurements. Comput. Netw. **17**, 107294 (2020)
32. Ricciato, F.: Traffic monitoring and analysis for the optimization of a 3G network. IEEE Wirel. Commun. **13**(6), 42–49 (2006)
33. Demšar, J., et al.: Orange: data mining toolbox in Python. J. Mach. Learn. Res. **14**(1), 2349–2353 (2013)
34. Raida, V., Svoboda, P., Rupp, M.: Lightweight detection of tariff limits in cellular mobile networks. In 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1–7. (2018)
35. Dischinger, M., Mislove, A., Haeberlen, A., Gummadi, K. P.: Detecting bittorrent blocking. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, pp. 3–8. (2008)
36. Ravaioli, R.: Active inference of network neutrality. Université Nice Sophia Antipolis. (2016)
37. Bifet, A., Gavalda, R.: Learning from time-changing data with adaptive windowing. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 443–448. (2007)
38. Meinshausen, N.: Quantile Regression Forests. J. Mach. Learn. Res. **7**, 983–999 (2006)

Veljko Pejović received his PhD in computer science from the University of California, Santa Barbara, USA in 2012 on the topic of resource-efficient wireless communication for rural areas. He is an assistant professor of computer science at the University of Ljubljana, Slovenia, where he works on mobile sensing and resource-efficient mobile computing.