INTERNET POLICY REVIEW
Journal on internet regulation

Volume 13 **Issue** 2

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

Mitigating information asymmetry in 5G networks

Hermann Bergmann Garcia e Silva
Faculty of Engineering of the University of Porto (FEUP)
Rúben Manuel Nunes Santos
Faculty of Engineering of the University of Porto (FEUP)
Manuel Ricardo *Faculty of Engineering of the University of Porto (FEUP)*

DOI: https://doi.org/10.14763/2024.2.1765

Published: 27 May 2024
Received: 31 October 2023 **Accepted:** 15 May 2024

Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Citation: Garcia e Silva, H.B., Santos, R.M.N., & Ricardo, M. (2024). Mitigating information asymmetry in 5G networks. *Internet Policy Review*, *13*(2). https://doi.org/10.14763/2024.2.1765

Keywords: Internet governance, Net neutrality, 5G, Traffic differentiation, NWDAF

Abstract: The implementation of traffic differentiation measures by internet service providers (ISPs) has raised concerns regarding net neutrality, potentially leading to discriminatory practices that challenge existing regulatory frameworks. The complexity of this issue intensifies with the advent of 5G networks as they dynamically assemble elements of the physical infrastructure to create logically segregated domains customised to accommodate usage scenarios with specific requirements, resulting in the categorisation of users, applications, and services into distinct groups which possess the capacity to disrupt the non-discriminatory treatment of data flows. Within this context, a pivotal question arises: how can regulatory authorities effectively evaluate traffic differentiation in 5G networks? In response, this paper proposes an innovative application of the standardised network data analytics function (NWDAF) to facilitate the assessment of internet traffic differentiation. We introduce this novel concept and demonstrate its implementation through a proof-of-concept prototype. By leveraging the NWDAF, regulators may obtain direct and automatic access to performance metrics of 5G networks, enabling the analysis of the traffic management mechanisms employed by ISPs.

Introduction

The mobile communications ecosystem has undergone a significant transformation in recent decades. It began in the early 1980s with the introduction of the first generation of cellular networks, primarily transmitting analog voice traffic as an extension of the public switched telephone network (PSTN). Since then, a new mobile network architecture has emerged, offering high-capacity broadband, low latency, and massive connectivity between individuals, objects, and sensors.

The advent of 5G marks a departure from the incremental improvements of previous generations. It introduces a customizable virtualized network capable of accommodating a wide range of applications and services while providing an unprecedented degree of flexibility and scalability in the orchestration of network resources. At the centre of 5G technology lies the innovative concept of network slicing, functioning as a mechanism intended to allocate and distribute network resources through virtualization. This approach assures end-to-end quality of service (QoS) and enables each logical network instance to support a unique service level agreement (SLA), providing user connectivity as seamless as if they were using a dedicated physical infrastructure.

Nonetheless, this versatile architecture framework establishes virtual domains for differential treatment of internet traffic. This capability raises concerns about upholding the net neutrality principle, which embodies the ideal of a platform that guarantees open and non-discriminatory access to information, safeguarding the internet from becoming a walled garden where internet service providers (ISPs) dictate which content, applications, and services are permitted and which are not. In the contemporary context, where traffic differentiation is of paramount importance, net neutrality must transcend the traditional notion of a network as a mere conduit and achieve a balance with QoS, allowing the categorisation of internet traffic into classes, as long as it is not discriminatory.

In this scenario, regulatory authorities face the challenge of assessing internet traffic differentiation in the dynamic 5G environment to identify potential practices that could undermine the net neutrality principle. This paper presents an approach that leverages the network data analytics function (NWDAF) to tackle the prevalent issue of high information asymmetry in telecommunications markets. This approach is designed to equip regulatory bodies with the capability to access real-time data directly from the 5G network, including critical performance metrics such as packet loss and packet delay.

As its primary contribution, this study introduces an innovative application of the standardised NWDAF, designed to facilitate the assessment of internet traffic management practices (ITMPs) by allowing controlled information exposure to regulatory authorities. Through the utilisation of the NWDAF, regulators gain automated access to the 5G network, enabling the evaluation of traffic differentiation mechanisms implemented by ISPs and enhancing transparency in 5G network operations.

The paper is organised into six sections. The first section delivers an overview of internet traffic discrimination, including practices such as blocking, filtering, throttling, paid prioritisation, and zero-rating. The second section examines the 5G network and its architecture. In the third section, the concept of the NWDAF is introduced. Subsequently, the fourth section focuses on assessing traffic differentiation in 5G networks. The fifth section describes the proof-of-concept design and the conducted tests. Finally, the last section provides concluding remarks.

Exploring internet traffic discrimination

Decisions related to network performance faced by ISPs are closely associated with two fundamental aspects: capacity management and traffic management. Capacity management encompasses a diverse range of choices concerning the network infrastructure, which includes physical links and equipment. These choices include server capacities, bandwidth allocation, and hardware upgrades, ensuring that the network can adapt to the ever-evolving demands of its users. Conversely, traffic management in packet networks comprises diverse mechanisms, such as resource reservation, traffic shaping, buffer management, and packet scheduling, governing the network's response to service requests which can be applied to specific network elements or extended across a network domain. They are organised into three logical planes: the data plane, responsible for the forwarding of data flows; the control plane, in charge of path selection; and the management plane, overseeing operational and administrative aspects related to user data traffic.

These mechanisms do not operate in isolation; instead, they interact in various combinations to manage and control internet traffic, ensuring network performance and meeting the QoS requirements of heterogeneous applications. However, it is important to note that while some of these measures represent acceptable network management practices that involve traffic differentiation, others may not, since they implement non-neutral data transmission. The criteria for making this distinction should be addressed in regulatory frameworks established by different jurisdictions, preserving the internet as an open, public network.

It is essential to acknowledge that in a regulatory framework upholding net neutrality, ISPs should only be allowed to engage in discriminatory practices under exceptional circumstances, notably to preserve network integrity or in congestion resolution. In such instances, these practices must be executed reasonably and to the overall benefit of the network and its users.

Therefore, the discriminatory treatment of data packets is associated with the idea of ISPs engaging in unacceptable traffic management practices. Nevertheless, the systematic categorisation of these practices has received limited attention from the research community, with most efforts directed toward case-by-case analysis. Few initiatives have been undertaken to explore and establish a common understanding of these discriminatory behaviours.

One of the most comprehensive studies in this field was conducted by Garret et al. (2018), introducing a taxonomy designed to consolidate terminology related to ITMPs. These practices were categorised based on the mechanisms employed by ISPs, the impact of these mechanisms on network traffic, and how users perceive such discrimination. The authors proposed four categories of differentiation mechanisms: block, delay, drop, and modify, which affect users' perception in different ways: blocked traffic, longer delays, increased jitter, throttling, and integrity violation.

Dischinger et al. (2010) associated discriminatory practices with traffic manipulation mechanisms. The authors consider that ISPs have various ways to treat classes of packets differently, including blocking, which involves terminating a data flow; deprioritizing, where ISPs may utilise multiple priority queues in routers to forward packets, allowing them to assign specific flows to lower-priority queues; packet dropping, where ISPs can selectively drop packets from a flow; modifying the Transmission Control Protocol (TCP) advertised window size, prompting the sender to slow down its data transmission; and application-level mechanisms, such as ISPs exerting control over an application's behaviour by modifying its protocol messages.

Jordan and Ghosh (2010) propose a framework to classify traffic management practices as reasonable or unreasonable based on the answer to four questions: where in the network, and at which layer, is the traffic management technique applied? What type of traffic management functionality is applied? Who decides whether the traffic management practice is applied? On what basis is it decided to apply the traffic management practice? The first question identifies whether the traffic management practice is used above or below the transport layer and whether it is

at an endpoint or a transit node. The second question seeks to characterise whether the practice is blocking, termination of a session, enhancement, or degradation of QoS. The third question is related to whether the decision has been made directly by the user or independently by the ISP. The last question aims to verify if the practice is applied to specific data traffic based on the application, source, destination, service provider, or payment.

Building on these previous works, this paper presents a categorisation of internet traffic discrimination and its associated practices, classifying them into distinct types that highlight actions by ISPs that contradict the idea of a non-discriminatory network which could justify regulatory intervention. These categories encompass three types: access discrimination, QoS discrimination, and price discrimination, as shown in Table 1. Access discrimination refers to the complete or partial restriction of accessing specific legal content, applications, or services on the internet (e.g. blocking, filtering). QoS discrimination impacts the perceived quality by enhancing or degrading the network performance of a specific service, application, or class of application (e.g. throttling, paid prioritisation). Price discrimination is related to increasing or decreasing the network access cost to a particular service, application, or class of application (e.g. zero-rating).

TABLE 1: Discriminatory practices

TYPES OF DISCRIMINATION	DESCRIPTION	PRACTICES
Access	Complete or partial restriction of accessing specific legal content, service, application, or class of applications	Blocking, Filtering
QoS	Enhancing or degrading the network performance of a specific service, application, or class of applications	Throttling, Paid prioritisation
Price	Increasing or decreasing the network access cost to a particular service, application, or class of applications	Zero-rating

Blocking

Blocking refers to the practice that prevents internet users from accessing specific applications, services, or content. Despite its historical roots dating back to the commercial utilisation of the internet in the 1990s, aiming to curb the proliferation of spam through email, blocking continues to persist as a contemporary phenomenon.

It can be implemented at various network infrastructure control points, including routers and gateways, operating at the individual, organisational, ISP, and national

levels. For example, at the individual level, users may employ parental control rules to restrict access to age-inappropriate content. Organisations can configure corporate firewalls to enforce access restrictions on websites that do not align with their internal policies. ISPs can establish blocking to mitigate threats like distributed denial of service (DDoS) attacks. Additionally, at the national level, the state may institute legal mandates for blocking specific digital content through a national gateway or impose such requirements on all ISPs operating in the country.

The practice of blocking digital content, applications, or services by ISPs can be segmented into four different methods, defined by the Office of Communications as primary techniques (Ofcom, 2011), which include: Internet Protocol (IP)-based, Uniform Resource Locator (URL)-based, Domain Name System (DNS)-based, and deep packet inspection (DPI) based blocking.

IP-based blocking impedes attempts to establish a connection between a device and a specific IP address or a defined set of addresses, preventing access to content hosted on servers in the network, which may include multiple websites or services. Another variation to IP-based blocking involves restricting access through port numbers embedded in the segment by transport protocols such as TCP or User Datagram Protocol (UDP).

Blocking based on a URL is executed by adding the URL to a denylist maintained by the ISP, preventing the connection to the requested server. DNS-based blocking involves controlling queries to the DNS database, intending to prevent or modify the responses to these queries. When the DNS server hosted by the ISP receives a query from the user's terminal equipment to resolve an IP address corresponding to a domain name, it can respond with a deliberate message, such as “IP address not found”, display a blocked page, or redirect to an IP address that does not correspond to the requested domain name.

Deep packet inspection (DPI) is a traffic management technology that enables ISPs to analyse the content of data packets in real-time. This capability enables network operators to handle traffic based on rules established at control points throughout the network, extending beyond the information contained in the packet header. Consequently, ISPs may employ DPI systems to identify, classify, redirect, assign different transmission priorities, or block packets containing specific payload content.

Several cases have been registered in which access to websites, applications, and

services was blocked. For example, in Brazil in 2004, broadband users of Brasil Telecom complained that the operator was selectively blocking the VoIP service provided by Skype. In 2005, Telus, one of Canada's largest ISPs, unilaterally blocked access to a server hosting the website Voices for Change, which supported the telecommunications workers' union in their efforts to raise awareness about a contentious labour dispute between the union and the company. In 2007, a major U.S. network provider, AT&T, collaborated with Apple to block all iPhone VoIP applications using its cellular network. In 2011, in the United States, AT&T, Verizon Wireless, and T-Mobile blocked the mobile payment application Google Wallet, hindering the service's access to subscribers and competing with their emerging application called ISIS (van Schewick, 2014). In 2012, Verizon was fined $1.25 million by the Federal Communications Commission (FCC) for blocking third-party tethering[1] applications on Android phones. More recently, in 2021, a regional ISP in the state of Idaho informed its users that access to Facebook and Twitter would be blocked by default in response to the decision of these social media platforms to suspend former President Donald Trump's account.

Filtering

The practice of filtering is typically considered to be equivalent to blocking, as it involves limiting access to undesirable or inappropriate online content without the knowledge or consent of affected users. While blocking entails a complete prohibition on accessing internet content or services, filtering represents a more nuanced form of impediment. Barnes et al. (2016) elucidate that the distinction between these two practices hinges on the matter of scale and perspective. Accordingly, filtering can be defined as a practice that restricts the flow of information on the internet by selectively scanning keywords, phrases, images, or strings, among others, performed on webpages, emails, social networks, chat rooms, newsgroups, electronic messages, video streaming, or executable files.

Similar to blocking, filtering is implemented at network control points, which may operate at various levels: individual, organisational, ISP, and national. It is important to note that internet filtering differs from filtering carried out by internet services. ISPs can perform filtering in the access network, internet exchange points, or other parts of the network infrastructure. In opposition, content and application providers (CAPs) employ filtering as part of their internal policies, which are integrated into their terms of use.

1. Tethering refers to using a mobile device as a modem to connect a separate device.

Filtering can be categorised into three different types: inclusion filtering, exclusion filtering, and content analysis, which can be utilised in combination to achieve the intended purpose. In common terms, inclusion filtering is often referred to as an “allowlist" since it permits users access to content that has been previously approved. On the other hand, exclusion filtering, the second type, involves the creation of a “denylist" to block undesirable content. The third type, content analysis, operates without predefined lists and focuses on analysing the requested content before granting user access (Hamilton, 2004; Eneman, 2010).

Rosenberg (2001) describes an additional filtering approach that involves the establishment of rating systems, which assign value judgments to categorise content along various dimensions, such as violence, nudity, language, sexual explicitness, and drug use.

One of the technologies employed for filtering is DPI, which enables ISPs to make decisions on how to manage traffic on the network. Hall et al. (2022) highlight that DPI typically involves using a separate copy of the data for analysis, keeping the original data flowing and preventing a significant impact on the network QoS. As noted by Parsons (2013), DPI's capability to inspect both the header and payload of datagrams grants this technology notable control over users' digital communications on the internet.

Fuchs (2012) argues that DPI can also be exploited for political control or social repression targeting specific groups in society. This is possible through keyword-based filtering, storage, and analysis, which enables the monitoring of individual users or groups, as well as the identification of the sender and receiver of the communication and the content transmitted.

In essence, internet filtering is rooted in the premise that certain content is harmful to society, necessitating the protection of citizens from exposure to materials that include child pornography, terrorism, racism, hate speech, intellectual property infringement, violence, and illegal gambling. However, opposing viewpoints sustain that internet filtering jeopardises fundamental human rights, most notably freedom of expression, and is viewed as a form of censorship that compromises the openness of the internet, undermining the foundations of democracies, which are constructed upon the rule of law (Enemann, 2010).

Throttling

Throttling refers to the deliberate act of degrading or delaying the transmission of internet traffic, thus reducing the transmission rate within the telecommunications

network. Conceived initially as a traffic differentiation measure to address network congestion, throttling restricts users' informational access when it no longer remains agnostic and becomes directed toward specific content providers, particular classes of applications, or users with a given data consumption profile such as heavy users.

Choffnes et al. (2017) argue that throttling is employed as a means to degrade the quality of services that compete with those offered by the ISP or to raise the barrier for new entrants in the market. Garret et al. (2018) also emphasise that the reason for this differentiated treatment is the pursuit of a competitive advantage, whereby an ISP prioritises its own services' traffic while degrading its competitors.

The ISP can implement throttling through a variety of techniques, including limiting the bandwidth via traffic shaping mechanisms that enforce a preconfigured rate by queuing excess packets (Li et al., 2019), carrying selected traffic over low-capacity or more congested links (Zhang et al., 2009), or applying distinct packet-forwarding priorities in routers (Lu et al., 2010).

In 2007, Comcast, a major cable ISP in the United States, employed DPI technology provided by Sandvine (specifically the Policy Traffic Switch Model 8210) to throttle the peer-to-peer file sharing application BitTorrent.

In 2011, the Canadian Games Association filed a complaint with the Canadian Radio-television and Telecommunications Commission (CRTC), alleging that ISP Rogers Communications Inc. was throttling popular online games such as World of Warcraft. An investigation carried out by the Compliance and Enforcement Sector confirmed that the ISP's technical ITPM significantly degraded the performance of such network traffic.

In 2017, the German regulatory authority, BNetzA, issued an administrative directive against Telekom Deutschland GmbH, prohibiting the use of throttling in its video streaming service, StreamOn, for the MagentaMobil tariffs, which were limited to a maximum of 1.7 megabits per second (Mbit/s), making it unviable to display video in high-definition quality.

Between January 2018 and January 2019, Li et al. (2019) conducted a one-year study in operational mobile networks that identified throttling in thirty ISPs in seven countries, mainly affecting video streaming services (e.g. YouTube, Netflix, and Vimeo).

Paid prioritisation

Paid prioritisation encompasses a practice characterised by granting privileged treatment to specific categories of traffic in exchange for financial compensation. This practice revolves around creating fast lanes within the network infrastructure, thereby establishing a hierarchical internet structure implemented through network management mechanisms.

According to the Center for Democracy and Technology (CDT), as described in a letter addressed to the FCC, paid prioritisation can be defined as the ISP practice of charging a fee to deliver the traffic of CAPs in an enhanced fashion over subscribers' last-mile facilities (Cooper et al., 2010). In this scenario, CAPs would pay a surcharge to leverage their QoS, while others would receive a best-effort service from the ISP.

Traditionally, the success or failure of new applications, services, and websites was primarily contingent on the quality of the content that was delivered and the perceived value attributed by users. However, with the possibility of ISPs establishing fast lanes, determining winners and losers in the online competition is influenced by those who can afford the costs associated with prioritising data traffic (van Schewick, 2015).

This logic creates an uneven playing field, where startups with innovative ideas may find themselves relegated to a slow lane, while large corporations with economic power enjoy the privilege of having their content transmitted seamlessly over the internet. This scenario threatens diversity and widespread access to information, undermining the principles that have made the internet an instrument for reducing communication barriers and empowering individuals.

In 2013, the French ISP Orange demanded an additional payment from Google in return for prioritising YouTube traffic on their mobile network infrastructure (Greenstein et al., 2016). The company CEO at the time openly acknowledged that the company's dominant presence in the promising African market played a pivotal role in facilitating this business arrangement, which was perceived as a way to compensate for the vast volume of Google traffic handled by the operator. While the specific financial details of the transaction were not disclosed, it symbolised the recognition of a payment made by a CAP to the ISP in exchange for preferential treatment of its data traffic.

Zero-rating

Zero-rating is a commercial practice employed by ISPs that offers free data traffic associated with specific content, applications, or services available on the internet, allowing end users to access content without the transmitted bytes being counted towards their monthly contracted allowance (data cap). It operates as a differentiated pricing model where the cost of using the telecommunications infrastructure is subsidised by the ISP or CAP (Garcia e Silva, 2017). Zero-rating is primarily feasible in service plans that limit the volume of data traffic users can access on the internet, making it more prevalent in the mobile telecommunications market due to bandwidth constraints (Marsden, 2016).

Gautier and Somogyi (2020) argue that the difference between zero-rating and paid prioritisation is that although both characterise discriminatory practices, the first is discrimination in terms of price, creating financial differentiation between CAPs, while the second is discrimination in terms of quality, attributing faster delivery to prioritised content.

Zero-rating offers can take different configurations, depending on the relationship between the ISP and CAP. As highlighted by Eisenach (2015), there are two more common types of practices. The first is the carrier-initiated zero-rating, in which the ISP provides free access to certain content as a strategy to attract new consumers and expand its market share. The second type is sponsored data, representing an arrangement where CAP subsidises the cost of accessing its content by compensating network operators to make it available to users at no data consumption cost.

Kak (2015) considers that zero-rating elevates content previously chosen by ISPs to privileged status, as they attribute a competitive advantage to the providers of this content. By making access less expensive and more attractive to users, these ISPs strengthen the market power of selected CAPs, often at the expense of excluding other competitors. Additionally, zero-rated content compromises the diversity of information and limits users' freedom of choice in the digital ecosystem.

Zero-rated business models include a broad spectrum, from individual websites like Wikipedia Zero to platforms featuring preselected applications. Notable examples include T-Mobile's Binge On and Free Basics by Facebook, the latter being the most extensive zero-rated program deployed in over sixty countries, with a primary presence in the African continent.

5G network

The 5G network comprises a suite of technologies intended to address the evolving demands of contemporary communication. These technologies are tailored to optimise traffic priorities and support at least three distinct usage scenarios with a unique set of performance requirements: enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (URLLC), and massive machine-type communications (mMTC). In order to accommodate these varying needs, the 5G architecture implements a partitioning of the physical infrastructure across access, transport, and core networks, enabling efficient resource allocation and network slice isolation, allowing each logical instance to operate independently while sharing the underlying physical infrastructure (Barakabitze et al., 2020).

Within the 5G core network, the user plane function (UPF) plays a crucial role as it acts as the interconnection interface between the radio access network (RAN) and the data network (DN). It performs essential functions such as packet routing, forwarding, inspection, downlink data buffering, QoS handling, and policy rules enforcement. The RAN is entrusted with providing radio access to user equipment (UE) and houses a key component known as the gNodeB (gNB). Beyond its primary radio resource management and session mobility responsibilities, the RAN is also responsible for QoS flow control and data routing to the UPF.

The session management function (SMF) oversees the UPF and manages the entire lifecycle of a protocol data unit (PDU) session[2], while the access and mobility management function (AMF) handles access authorization, authentication, UE location tracking, and connection management. Figure 1 illustrates the 5G architecture.

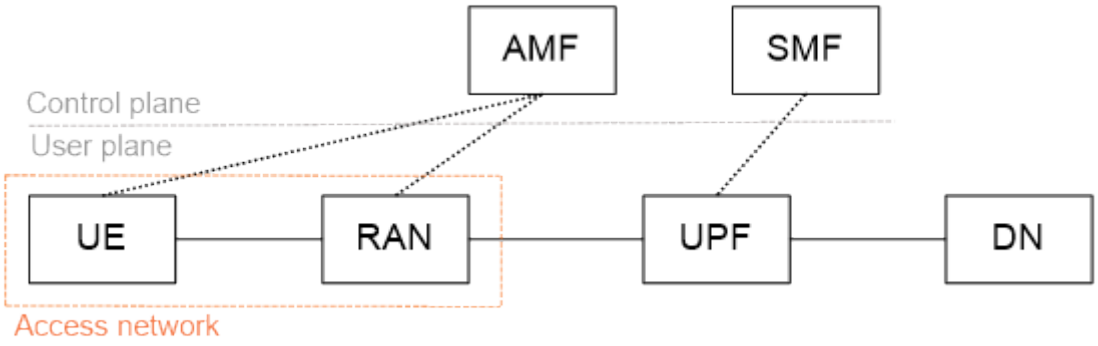

FIGURE 1: 5G network architecture.

2. PDU session represents a logical connection between the UE and the DN.

Network data analytics function (NWDAF)

The NWDAF is a standardised function of the 5G core network introduced by the 3rd Generation Partnership Project (3GPP)[3] in Release 15 (3GPP, 2018) and enhanced in subsequent releases. Its primary mission is to centralise data collection from other core network functions (NFs), application functions (AFs), and Operations, Administration, and Maintenance (OAM) systems. This functionality enables the NWDAF to provide statistical analysis and prediction capabilities for monitoring the user experience and network performance.

One of the distinguishing features of the NWDAF is its consumer-producer architecture, which is established through standard interfaces that enable the periodic collection of information related to network status and resource utilisation on a per-slice basis. The data collected is used to deliver insights into past events and predictive analysis in response to requests or subscriptions from other NFs.

As outlined in the technical specification TS 29.520 (3GPP, 2020), the NWDAF offers a comprehensive range of services, granting consumer NFs access to various categories of analytical data, including network performance metrics, QoS sustainability, user data congestion, and observed service experiences. These services are also accessible to external entities through the network exposure function (NEF), allowing communication with the 5G core.

Regarding performance measurements within 5G networks, mobile network operators can configure the NWDAF to invoke existing OAM services and retrieve relevant data to generate analytics. The data collection process from the OAM begins with the NWDAF subscribing to notifications related to the services provided by the corresponding management service producer. Subsequently, this producer responds to the NWDAF, indicating the success or failure of the subscription. In the following stages, the collected data is prepared, and the producer notifies NWDAF when the file is ready for retrieval. Finally, the NWDAF retrieves the data (3GPP, 2021).

Following the conclusion of the data collection phase, the NWDAF employs advanced analytics techniques to extrapolate valuable insights concerning network performance and QoS. This repertoire of techniques includes traffic analysis, anomaly detection, predictive modelling, and correlation of different network events. As a result, NWDAF can identify patterns and trends in real-time and his-

3. The 3GPP unites seven telecommunications standard organisations that develop specifications for mobiletelecommunications.

torical datasets.

These analytical capabilities are crucial in facilitating informed decision-making for network operators (Niu et al., 2022). Leveraging the data provided by the NWDAF, operators can optimise network performance and allocate resources more efficiently. These findings can also support regulators in assessing the alignment of traffic differentiation measures with obligations related to net neutrality. The outputs generated by the NWDAF can be presented in a variety of formats, including reports, dashboards, and graphical representations, to facilitate the interpretation of analytical findings.

Traffic differentiation in 5G networks

In assessing traffic differentiation within a dynamically assembled logical network model, considerable challenges arise due to the numerous possible configurations that affect network management. It is, therefore, unrealistic to expect a 5G network to uniformly accommodate slices designed for various purposes, such as time-critical communications, high data rates, or support for Internet of Things (IoT) applications. Recognizing this reality, a pragmatic approach must be adopted in interpreting the net neutrality principle.

To address this complexity, it becomes necessary to examine the non-discriminatory treatment of internet traffic in 5G networks on a per-slice basis, particularly for those slices that are made publicly accessible to users or CAPs. Monitoring net neutrality should be conducted per slice, refraining from direct comparisons among slices with distinct attributes.

The assessment of net neutrality can be carried out by capitalising on the usage scenarios in the 5G network, namely eMBB, URLLC, and mMTC. It is essential to acknowledge that each type of network slice places distinct emphasis on its performance requirements.

In the case of an eMBB slice, the allocation of throughput to traffic flows is of paramount significance, as it is tailored for multimedia applications. In contrast, for a URLLC slice, the focus is on achieving low latency and ensuring high reliability. Finally, for a mMTC slice, the primary consideration is connection density.

In this manner, the most relevant performance parameters of each 5G slice serve as the required inputs for assessing whether the traffic management mechanisms employed by mobile network operators may give rise to discriminatory practices.

The regulatory authority may demand the mobile network operator collect data on relevant metrics for each 5G public slice, leading to the establishment of a regulatory monitoring process reliant on NWDAF analytics outputs, as depicted in Figure 2. It is noteworthy that, in this context, the responsibility for securely exposing 5G data analytics to external parties, whether for requests or subscriptions, lies with the network exposure function (NEF).

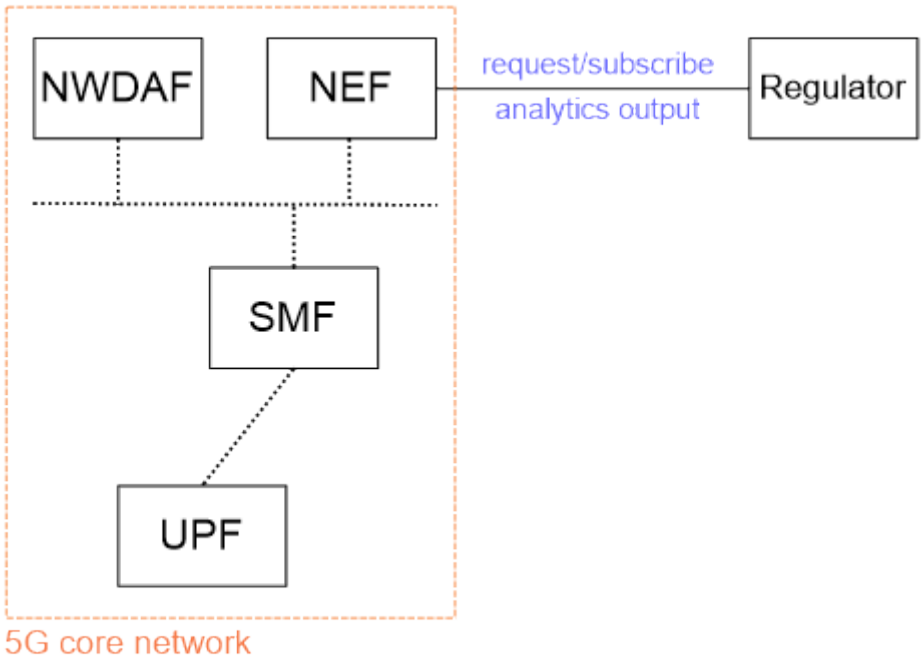

FIGURE 2: Regulatory access to performance data.

Once the regulatory framework is in place, two distinct approaches can be employed to monitor traffic differentiation in 5G slices. The first approach is preventive, involving regular and systematic monitoring to identify patterns that may suggest potential net neutrality violations. This proactive method also generates data to improve transparency regarding traffic management practices. In contrast, the second approach is reactive and is activated when a stakeholder files a complaint, culminating in case-specific monitoring.

Nevertheless, the complex task of detecting the root cause of traffic differentiation entails the examination of multiple variables, including congestion levels experienced by the service, justified cases of traffic differentiation (e.g. specialised services, illegal content or cyber-attacks), and the emergence of content delivery networks (CDNs).

In addressing traffic differentiation in 5G networks, a promising strategy involves the comparison of data packets from different flows that share common QoS at-

tributes. The objective is to determine whether these packets exhibit similar network performance. This evaluation requires the designation of a baseline CAP (bCAP) as a reference, which is compared with a target CAP (tCAP) to identify any potential traffic differentiation implemented by ISPs.

Among the network performance parameters delineated in standards and technical specifications, such as Y.1540 (ITU, 2011) and TS 28.552 (3GPP, 2022), packet loss[4] and packet delay[5] stand out as of particular significance. These metrics are directly affected by traffic differentiation mechanisms, making them important for understanding network behaviour.

Packet loss may reveal issues in cellular networks, including interference, insufficient signal strength, or network congestion. Besides, packet loss also provides insights into traffic degradation resulting from packet scheduling and buffer management policies.

In contrast, packet delay is influenced by several variables, including packet length, distance between communication peers, number of nodes traversed by data packets, and the level of network congestion. It is also instrumental in evaluating the dynamics of internal queuing delays, which may lead to discriminatory practices by ISPs.

The proposed strategy evaluates traffic differentiation by collecting and analysing packet loss and packet delay measurements along a specific network segment. To do so, monitoring points should be identified at different boundaries within the 5G network. The selection of the UPF and RAN as monitoring points is based on their critical roles in end-to-end communication, enabling the identification of actions that differentiate certain types of traffic based on predefined criteria. Excluding measurements in the RAN-to-UE segment helps isolate variables such as mobility and transmission over the air interface.

Therefore, the NWDAF can gather measurements from the UPF and RAN, enabling separate assessments of downlink (DL) and uplink (UP) traffic streams. This separation is necessary because each direction may exhibit potential variations in performance. A visual representation of this approach is illustrated in Figure 3.

4. Packet loss represents the packets that fail to reach their intended destination, including packets dropped, packets lost in transmission, and packets received in the wrong format.

5. Packet delay refers to the time required for a data packet to travel the distance between the sending and receiving endpoints, combining the processing, queueing, transmission, and propagation delays.

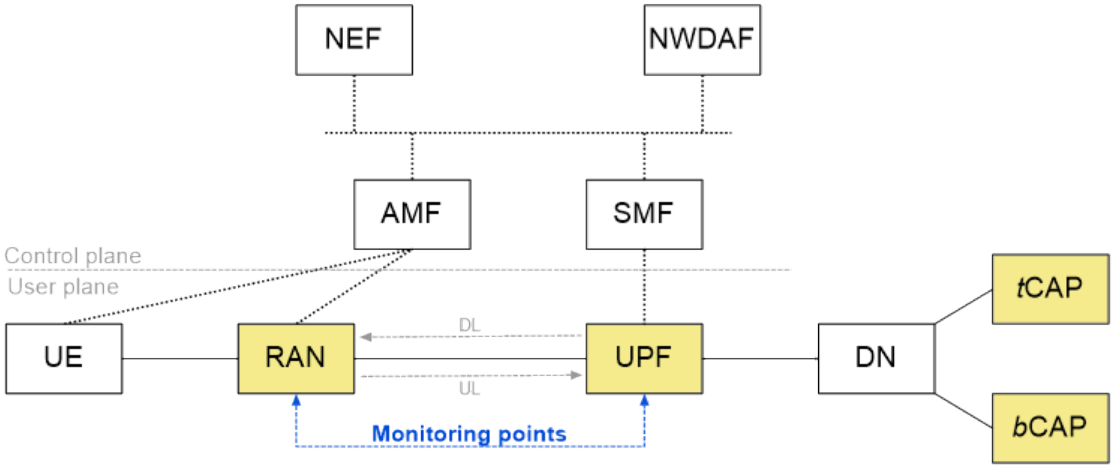

FIGURE 3: Comparing target CAP (tCAP) and baseline CAP (bCAP) traffic.

UPF can be deployed in a decentralised architecture with multiple instances closer to the edge network. This deployment aims to reduce the routing distance between a service and the users' locations. In this architecture, one user plane funcion acts as the PDU Session Anchor (PSA) for the other sequential instances. For example, an intermediary UPF (I-UPF) can be inserted as a ramification point between two edge UPFs, as demonstrated in Figure 4. The existence of numerous possible configurations also justifies the necessity of monitoring points at the PSA UPF and RAN, as traffic differentiation may be implemented by ISPs across different UPFs along the data path.

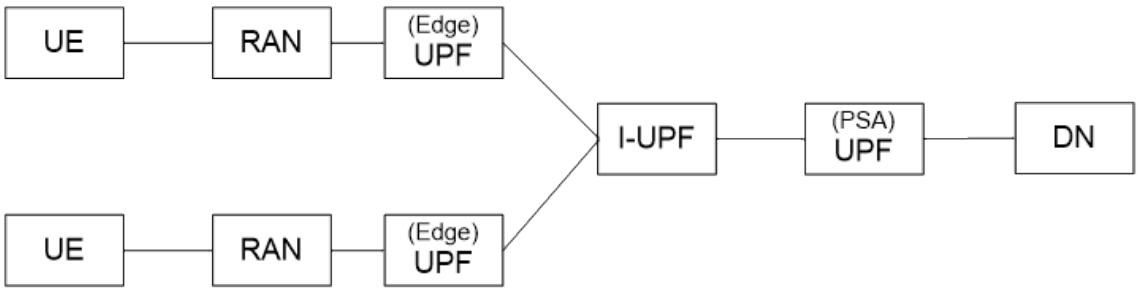

FIGURE 4: Deployment of multiple UPFs.

If the ISP abstains from implementing any traffic differentiation mechanism, it is expected that both CAPs (baseline and target) would exhibit similar levels of loss and delay, with negligible variations. This equivalence in outcomes signifies the fair treatment of traffic types. As a result, the QoS classification and priority level assigned to tCAP and bCAP packets should be identical, maintaining neutral packet scheduling and buffer management policies, which represents the null hypothesis.

However, statistically significant differences in measurements are likely to indicate traffic differentiation within the network segment. For example, the ISP might prioritise bCAP packets over tCAP packets, applying distinct packet scheduling and buffer management policies to each traffic class. Under such circumstances, higher loss rates may suggest using a buffer management policy that selectively drops incoming tCAP packets with a higher probability than bCAP packets (e.g. weighted random early detection — WRED). Likewise, prolonged delays may signal a packet scheduling policy that gives higher priority (e.g. strict priority — SP) or allocates a larger share of the bandwidth to bCAP packets (e.g. weighted fair queueing — WFQ). Another differentiation method involves limiting the transmission of tCAP packets using traffic shaping mechanisms at critical nodes, such as the UPF.

It must be emphasised that traffic differentiation does not inherently imply a violation of the net neutrality principle (van Schewick, 2015). Its implementation can serve legitimate purposes, such as ensuring network integrity and security or managing network congestion, demanding a holistic analysis by regulatory authorities. Nonetheless, the proposed approach illustrates how real-time 5G data analytics can enhance transparency in traffic management policies, enabling the regulator to access relevant performance data and monitor ITMPs within a net neutrality regime.

Proof-of-concept and test results

In order to evaluate the feasibility of employing the NWDAF as a strategic tool to enhance the transparency of traffic management practices and enable the acquisition of performance data related to the 5G network by regulatory authorities, a system was designed to establish a connection between a UE receiving two video streams from distinct CAPs over the internet.

Based on the defined premises, driven by the compelling need for expertise in programming languages like Python and JavaScript, the experiment was conducted by Santos (2023) as part of the Master's Program in Electrical and Computer Engineering at the Faculty of Engineering of the University of Porto (FEUP).

The eMBB usage scenario was chosen as a case study for its direct applicability to services that require high data rates and significant bandwidth, in which bCAP and tCAP are represented by web pages that host multimedia content, including platforms such as YouTube, Radio and Television of Portugal (RTP), Portuguese Independent Television (TVI), and Portuguese Independent Communication Society (SIC), all simultaneously accessed by the UE. This approach enables a comparative

assessment of video streams, specifically focusing on performance metrics such as packet loss and delay. As depicted in Figure 5, references marked as “t1” and “t2” are used to designate monitoring points in the data flow. More precisely, “t1” represents both the point and moment when a packet enters the UPF queue, while “t2” represents its departure for routing towards the corresponding gNB. It is noteworthy that, as an external entity, the regulatory authority has the technical capability to access this data through the network exposure function (NEF).

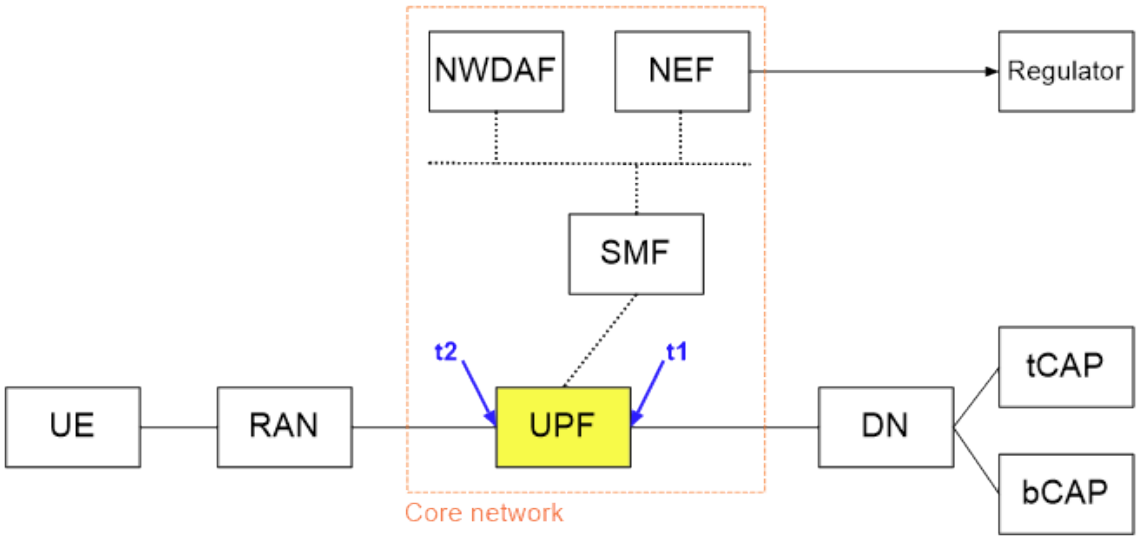

FIGURE 5: Proof of concept.

This configuration was established by employing the 5G core network to capture data flows at the UPF and obtain key performance indicators (KPIs) through the NWDAF. The emulation of the 5G core network was implemented on a host PC using the open-source software Open Air Interface (OAI), while the UE and RAN were simulated using the UERANSIM[6] application. OAI was chosen for modelling the 5G core network due to its extensive documentation and compatibility with UERANSIM.

OAI utilises Docker container[7] technology to deploy various network nodes within the 5G core network, offering scalability, portability, and isolation of network functions. Since the primary focus did not involve configuring radio components in the 5G network, UERANSIM was selected to simultaneously simulate the UE and the

6. Open source 5G UE and RAN simulator that is used for testing the 5G core network and studying the 5G system.

7. A container provides a lightweight runtime environment that isolates packages, applications, and their dependencies which operate independently, ensuring consistent performance regardless of the host environment. A Docker container is a specific type of container that utilises the Docker platform.

RAN. The UERANSIM is encapsulated in a Docker container, simplifying deployment and enabling connections to video streams provided by the CAPs.

The integration of NWDAF involved adding a container to the network, expanding the range of network functions provided by the OAI. NWDAF functions as a host and includes a web server responsible for collecting data from the UPF.

The CAPs are structured as individual web pages, each hosting a unique video stream. These streams run simultaneously, enabling a comparative analysis of performance metrics. Furthermore, deliberate interference is introduced by the ISP into one of the streams (tCAP) to influence its performance dynamics in a controlled manner for assessment.

In practical implementation scenarios, the regulatory authority typically acts as an internet node positioned outside the 5G core network. However, in the context of a self-contained experiment the host PC takes on the role of the regulator. This choice aligns with the primary objective of demonstrating the proper operation of the developed NWDAF, integrated into the NFs deployed by the OAI as another container in the network rather than replicating an external entity precisely.

To assess the impact of an ISP on data flows, traffic manipulation was conducted using the "tc" (traffic control) tool on Ubuntu[8]. This tool enables the creation of a distinct queue in a specified interface, facilitating the differential handling of incoming traffic directed toward the UE. The configuration of traffic profiles for this queue requires the implementation of a filtering mechanism to identify specific traffic flows for manipulation.

After creating the queue and implementing traffic filtering, the "tc" tool is used for traffic shaping, which involves introducing delays and packet loss into the network traffic. Depending on user preferences, these delay and loss parameters can be applied individually or simultaneously. Moreover, the "tc" tool provides the flexibility to fine-tune delay and loss settings without removing the previously instantiated queue, enhancing the traffic manipulation setup's adaptability.

In order to replicate the ISP intervention, a deliberate introduction of 20% packet loss and a 50 milliseconds (ms) delay were applied to the tCAP video flow stream. The primary objective was to ascertain the capability of the NWDAF in identifying occurrences in both metrics, comprising two distinct scenarios. The first scenario pertains to HTTP/2[9] traffic, while the subsequent scenario is associated with

8. Open-source Linux operating system.

HTTP/3[10]. The outcomes of these test scenarios are discussed below.

i) HTTP/2: TVI was designated as the bCAP (Flow 1), and SIC was assigned as the tCAP (Flow 2). The results show that SIC packets experienced a delay of approximately 50 milliseconds and a packet loss ratio of 17.72%, consistent with the predetermined experimental conditions. Furthermore, it was also observed that the TVI stream's average delay and packet loss ratio remained at low levels, indicating the absence of interference by the ISP, as depicted in Figure 6.

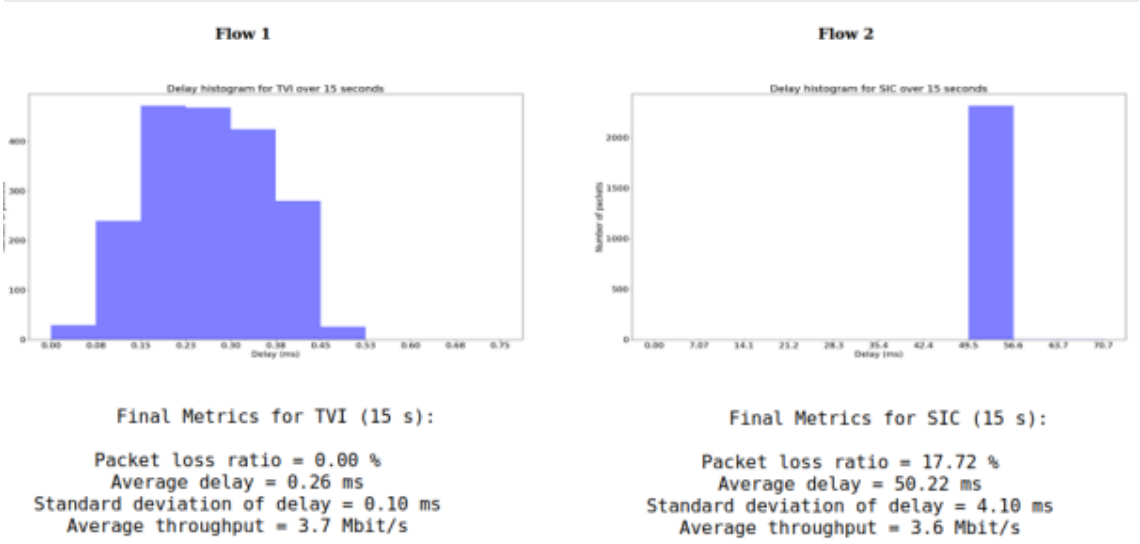

FIGURE 6: Delay and packet loss ratio for HTTP/2 (Santos, 2023, p. 59).

ii) HTTP/3: RTP was designated as the tCAP (Flow 1), and YouTube was assigned as the bCAP (Flow 2). The results show that RTP packets experienced a delay of approximately 50 milliseconds and a packet loss ratio of 23.69%, consistent with the predetermined experimental conditions. It was also observed that the YouTube stream's average delay and packet loss ratio remained at low levels, indicating the absence of interference by the ISP, as illustrated in Figure 7.

9. HTTP/2 is a major revision of the Hypertext Transfer Protocol (HTTP) designed to improve the speed and efficiency of web communication that uses the TCP as its transport layer.
10. HTTP3 is the latest version of Hypertext Transfer Protocol (HTTP) designed to enhance the speed, security, and efficiency of web communication that uses the QUIC (Quick UDP Internet Connections) as its transport layer.

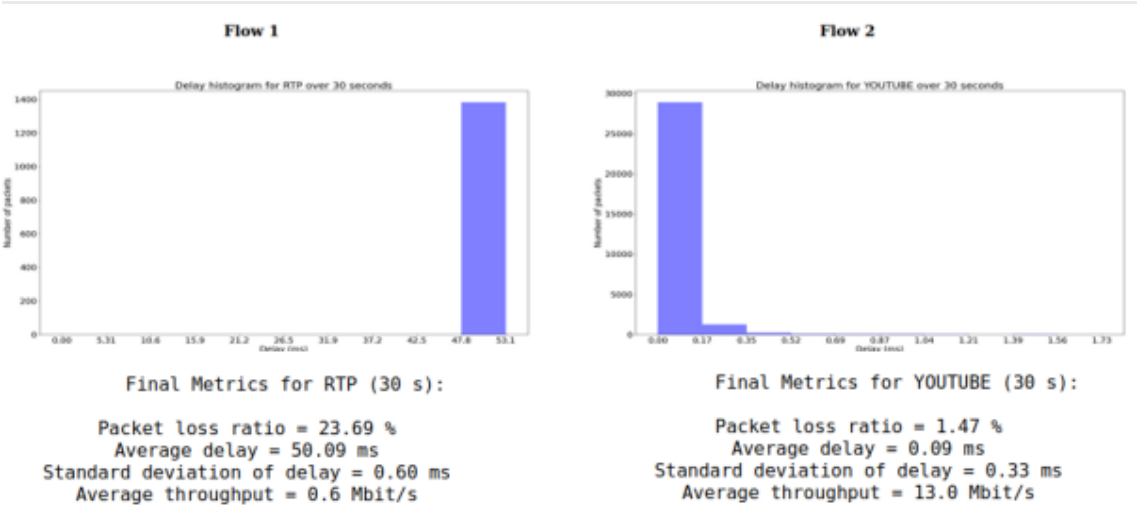

FIGURE 7: Delay and packet loss ratio for HTTP/3 scenario (Santos, 2023, p. 59).

The series of tests facilitated the assessment of the system's functionality and validated the NWDAF function. This validation relied on its ability to extract metrics from the traffic captured at the UPF and the results demonstrate the NWDAF's potential in deriving performance metrics for evaluating traffic differentiation in 5G networks. As a result, the NWDAF acts as a bridge between mobile operators and regulators, providing direct and automated access to network performance data, thereby reducing information asymmetry in addressing ISPs' discriminatory practices.

Conclusion

This research has addressed the challenges of upholding the non-discriminatory treatment of internet traffic in the dynamic 5G environment. In particular, the study made an overview of internet traffic discrimination, analysed ISPs discriminatory practices, examined the 5G network and its architecture with particular focus on the NWDAF, and explored traffic differentiation within a programmable mobile network architecture.

The categorisation of internet traffic discrimination into access, QoS, and price discrimination provides a structured reference to distinguish the ISP practices that affect net neutrality. By aligning each practice with its respective category, this approach contributes to a clearer comprehension of the ISPs' actions that contradict the idea of an open and equitable network, making it possible to qualify them as unacceptable ITMPs.

This paper also presents a novel approach centred on the NWDAF for assessing

traffic differentiation in 5G networks. This approach equips regulatory authorities with the technical capability to access essential performance metrics, supporting the evaluation of network management practices implemented by ISPs and providing a valuable method to improve regulatory oversight.

The proof-of-concept demonstrates the NWDAF's ability to collect performance metrics in an eMBB scenario. This demonstration provides insights into the traffic differentiation mechanisms employed by ISPs. Therefore, the assessment of net neutrality in the context of 5G networks involves the establishment of a regulatory monitoring process that relies on the analytical outputs of the NWDAF. Nevertheless, in real-world applications, ISPs may impose resistance on regulatory authorities' attempts to access the 5G core network and retrieve the NWDAF's data. This potential implementation challenge underscores the need for regulatory action grounded in a specific legal framework, which varies depending on the jurisdiction.

In future research initiatives, it is recommended to adopt a similar approach to include other monitoring points in the 5G network and overcome the UPF-centric model. Another necessity is to explore other usage scenarios, such as URLLC and mMTC, and promote a holistic understanding of the 5G architecture. There is also a demand to design an interface for regulatory authorities to request and retrieve performance data from the NWDAF. From a quantitative perspective, these investigations should also include an in-depth examination of the representative metrics of each 5G usage scenario, aiming to identify patterns that could indicate discriminatory practices by ISPs.

References

3GPP. (2018). *System architecture for the 5G System (5GS)* (Technical Specification (TS) 23.501). https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

3GPP. (2020). *5G System; Network data analytics services; Stage 3* (Technical Specification (TS) 29.520). https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3355

3GPP. (2021). *Architecture enhancements for 5G System (5GS) to support network data analytics services* (Technical Specification (TS) 23.288 v17.0.0). https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3579

3GPP. (2022). *Management and orchestration; 5G performance measurements* (Technical Specification (TS) 28.552 v18.0.0). https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3413

Barakabitze, A. A., Ahmad, A., Mijumbi, R., & Hines, A. (2020). 5G network slicing using SDN and

NFV: A survey of taxonomy, architectures and future challenges. *Computer Networks*, *167*, Article 106984. https://doi.org/10.1016/j.comnet.2019.106984

Barnes, R., Cooper, A., Kolkman, O., Thaler, D., & Nordmark, E. (2016). *Technical considerations for internet service blocking and filtering* (Report RFC7754). Internet Architecture Board. https://doi.org/10.17487/RFC7754

Choffnes, D., G., P., & M, A. (2017). *An empirical evaluation of deployed DPI middleboxes and their implications for policymakers*. TPRC45: Research Conference on Communications, Information and Internet Policy. https://people.cs.umass.edu/~phillipa/papers/DPI.pdf

Cooper, A., Morris, J., & Sohn, D. (2010, September 8). *Re: Preserving the open internet, GN docket no. 09-191; Framework for broadband internet service, GN docket no. 10-127* [Letter to M. Dortch]. https://cdt.org/wp-content/uploads/pdfs/CDT_FCC%20Letter_9_8_10.pdf

Dischinger, M., Macron, M., Guha, S., Gummadi, K. P., Mahajan, R., & Saroiu, S. (2010, April 28). Glasnost: Enabling end users to detect traffic differentiation. *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*. NSDI'10. https://dl.acm.org/doi/10.5555/1855711.1855738

Eisenach, J. A. (2015). *The economics of zero rating* [Study]. NERA Economic Consulting. https://www.nera.com/insights/publications/2015/the-economics-of-zero-rating.html

Eneman, M. (2010). Internet service provider (ISP) filtering of child-abusive material: A critical reflection of its effectiveness. *Journal of Sexual Aggression*, *16*(2), 223–235. https://doi.org/10.1080/13552601003760014

Fuchs, C. (2012). *Implications of Deep Packet Inspection (DPI) internet surveillance for society* [Technical report]. https://westminsterresearch.westminster.ac.uk/item/96vwx/implications-of-deep-packet-inspection-dpi-internet-surveillance-for-society

Garcia e Silva, H. B. (2017). *Neutralidade de rede: A prática do zero-rating e o Marco Civil da Internet [Network neutrality: The practice of zero-rating and the Internet Civil Framework]* [Master's thesis, FUMEC University]. https://repositorio.fumec.br/xmlui/handle/123456789/412

Gautier, A., & Somogyi, R. (2020). Prioritization vs zero-rating: Discrimination on the internet. *International Journal of Industrial Organization*, *73*, Article 102662. https://doi.org/10.1016/j.ijindorg.2020.102662

Greenstein, S., Peitz, M., & Valletti, T. (2016). Net neutrality: A fast lane to understanding the trade-offs. *Journal of Economic Perspectives*, *30*(2), 127–150. https://doi.org/10.1257/jep.30.2.127

Hall, J. L., Aaron, M. D., Adams, S., Andersdotter, A., Jones, B., & Feamster, N. (2020). *A survey of worldwide censorship techniques* (Internet-Draft 04). Internet Research Task Force. https://datatracker.ietf.org/doc/html/draft-irtf-pearg-censorship-04

Hamilton, S. (2004). *To what extent can libraries ensure free, equal and unhampered access to internet-accessible information resources from a global perspective?* [Doctoral thesis, University of Copenhagen]. https://doi.org/10.31237/osf.io/dzmvr

International Telecommunication Union. (2011). *Network performance objectives for IP-based services* (Recommendation Y.1541; Itu-T Series Y). https://www.itu.int/rec/T-REC-Y.1541-201112-I/en

Jordan, S., & Ghosh, A. (2010). A framework for classification of traffic management practices as reasonable or unreasonable. *ACM Transactions on Internet Technology*, *10*(3), 1–23. https://doi.org/10.1145/1852096.1852100

Kak, A. (2015). *The internet únbundled: Locating the user's voice in the debate on zero-rating* [Master's thesis]. University of Oxford.

Li, F., Niaki, A. A., Choffnes, D., Gill, P., & Mislove, A. (2019). A large-scale analysis of deployed traffic differentiation practices. *Proceedings of the ACM Special Interest Group on Data Communication*, 130–144. https://doi.org/10.1145/3341302.3342092

Lu, G., Chen, Y., Birrer, S., Bustamante, F. E., & Li, X. (2010). POPI: A user-level tool for inferring router packet forwarding priority. *IEEE/ACM Transactions on Networking*, *18*(1), 1–14. https://doi.org/10.1109/TNET.2009.2020799

Marsden, C. (2016). Comparative case studies in implementing net neutrality: A critical analysis of zero rating. *SCRIPTed*, *13*(1). https://doi.org/10.2966/scrip.130116.1

Niu, Y., Zhao, S., She, X., & Chen, P. (2022). *A survey of 3GPP release 18 on network data analytics function management*. 146–151. https://doi.org/10.1109/ICCCWorkshops55477.2022.9896472

Ofcom. (2011). *"Site Blocking" to reduce online copyright infringement: A review of sections 17 and 18 of the Digital Economy Act* [Report]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/78095/Ofcom_Site-Blocking-_report_with_redactions_vs2.pdf

Parsons, C. (2013). *The politics of deep packet inspection: What drives surveillance by internet service providers?* [Doctoral thesis, University of Victoria]. https://dspace.library.uvic.ca/server/api/core/bitstreams/5531407d-7778-466a-8cfc-ac47951e91bb/content

Rosenberg, R. S. (2001). Controlling access to the internet: The role of filtering. *Ethics and Information Technology*, *3*, 35–54. https://doi.org/10.1023/A:1011431908368

Santos, R. D. N. (2023). *Net neutrality in the 5G/6G era* [Master's thesis, University of Porto]. https://hdl.handle.net/10216/152168

van Schewick, B. (2014, May 6). The case for rebooting the network-neutrality debate. *The Atlantic*. http://www.theatlantic.com/technology/archive/2014/05/the-case-for-rebooting-the-network-neutrality-debate/361809/

van Schewick, B. (2015). *The case for meaningful network neutrality rules* [White paper]. https://cyberlaw.stanford.edu/publications/case-meaningful-network-neutrality-rules

Zhang, Y., Mao, Z. M., & Zhang, M. (2009). Detecting traffic differentiation in backbone ISPs with NetPolice. *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, 103–115. https://doi.org/10.1145/1644893.1644905

Published by
Alexander von Humboldt Institute for Internet and Society

in cooperation with
CREATe
cnrs centre internet et société
R&I IN3 Internet interdisciplinary Institute Universitat Oberta de Catalunya
University of Tartu Johan Skytte Institute of Political Studies