Computer Communications 127 (2018) 86–94

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Exploring DSCP modification pathologies in the internet☆

Ana Custura, Raffaello Secchi*, Gorry Fairhurst

University of Aberdeen, Fraser Noble Building, Aberdeen AB24 3UE, UK

A B S T R A C T

The differentiated service architecture (DiffServ) provides a means for network devices to classify traffic based on the DiffServ codepoint (DSCP) and to map the traffic to a specific QoS forwarding treatment. Successful use beyond the local network depends on consistent remarking and forwarding of the DSCP value inside and at the boundaries of DiffServ domains. This paper provides the results of a new widescale measurement campaign to examine how the DSCP value is altered as packets travel along a set of Internet paths. This allows us to infer whether a packet is likely to receive an appropriate QoS treatment and to comment the opportunities for more widely deploying DiffServ QoS. Our results identify a set of remarking pathologies, revealing that many deployed routers continue to use the previous semantics of the deprecated Type of Service (ToS) field. We also note that is not common to observe clearing of the bits in the DiffServ field, as previously believed for routers in the core of the Internet, although this varies significantly depending on the type of network studied. Our results are related to recent IETF work that recommends use of specific DSCP values.

1. Introduction

Differentiated Services (DiffServ) was first introduced in 1998 to allow different Internet flows to receive a specified Quality of Service (QoS) treatment [1]. It was introduced to overcome the scalability and complexity limitations of more fine-grained architectures, by providing consistent treatment of traffic with similar QoS requirements (a traffic class) as this is forwarded along a path. In almost 20 years of service, DiffServ has been integrated with all the principal network technologies, is largely implemented in access and core networks, and has been implemented in all major operating systems. However, DiffServ has not emerged as the final answer to the problem of bringing QoS to the Internet and its current use is mainly limited to network operator domains.

While DiffServ can readily be deployed within a DiffServ domain, there are also opportunities for using DiffServ across multiple domains along an Internet path. This paper discusses the requirements for realising end-to-end use of DiffServ and the barriers for this deployment. Supported by an analysis of Differentiated Services Code Point (DSCP) probes over a range of Internet paths, we show how the misalignment of practices in processing the DiffServ field in different autonomous systems (AS), remarking caused by deprecated DSCP semantics still in use in old equipment, and incorrect router configurations still have serious implications on the ability to use DiffServ end-to-end. As an example, we found remarking of DSCPs on certain paths resulted in *priority inversion*, i.e. priorities indicated by the DiffServ field were exchanged after processing in a sequence of nodes, which could lead to severe disruption of priority traffic.

The debate on interconnection of DiffServ domains and harmonisation of practices was recently re-ignited at the Internet Engineering Task Force (IETF) with the publication of Geib and Black [2]. This demonstrates the interest in maintaining DSCP markings across interconnected domains and proposes a small number of interconnection classes on MPLS (Multi-protocol Label Switching) networks to support inter-operation.

Preserving consistent DSCP markings is a critical pre-requisite for end-to-end QoS across the Internet. Within the Internet core, the DSCP is usually forwarded transparently. Even in over-provisioned core networks, DiffServ PHBs can improve robustness of the service (e.g., to mitigate the impact of DDoS traffic [3]).

Whereas in some edge and datacenter networks a range of remarking pathologies can be observed, some networks reset the DSCP value of every packet at entry to their domain. This causes complete loss of class marking and precludes use of appropriate Per-Hop Behaviour (PHB)s further along the path. Our measurements show that these practices are not infrequent. Analysing a dataset of over 25,000 source-destination pairs, we found over 70% of sampled paths showed some form of DSCP modification pathology.

☆ This work is funded by the European Unions Horizon 2020 research and innovation programme under grant agreement no. 644399 (MONROE) through the Open Call and grant agreement no. 644334 (NEAT). The views expressed are solely those of the author(s). The European Commission is not responsible for any use that may be made of that information.

* Corresponding author.

E-mail addresses: ana@erg.abdn.ac.uk (A. Custura), r.secchi@abdn.ac.uk (R. Secchi), gorry@erg.abdn.ac.uk (G. Fairhurst).

https://doi.org/10.1016/j.comcom.2018.05.016

Received 12 December 2017; Received in revised form 18 April 2018; Accepted 28 May 2018

We argue that inconsistent modification is not only damaging to a proper functioning of DiffServ, but also jeopardises the ability to accommodate new traffic classes. One example is recent IETF work to specify a DSCP value for the Lower Effort (LE) service [4], which requires careful consideration to reduce the impact of remarking pathologies.

The remainder of the paper is organised as follows: Section 2 reviews use of DiffServ and related work. Section 3 describes our methodology, where we introduce a new tool, PathTrace, to analyse DSCP modification. This tool was used from vantage points across the Internet between December 2016 and July 2017. We present the results of our measurement in Section 4, these extend preliminary work in Custura et al. [5]. A discussion of the implications on deploying DiffServ QoS follows in Section 5. We also comment upon the implications on current IETF standards activities. Finally, the conclusion summarises our key findings.

2. Background

2.1. Differentiated services code points

Classification of Internet traffic was first supported by the 8-bit ToS field in the IPv4 header. The field was divided into two sub-fields: the three most significant bits assigned the packet to one of eight precedence classes, the five least significant bits specified traffic characteristics. The ToS field was later re-purposed to carry a DSCP [1] in the top 6 bits and the Explicit Congestion Notification (ECN) [6] field in the bottom 2 bits.

The 6 bit DSCP field makes available 64 codepoints. These have been divided into three groups by the Internet Assigned Number Authority (IANA)[7]. The 32 codepoints with a zero least significant bit (even codepoints) define standard DiffServ classes. The 16 codepoints with the two least significant bits “11” are available for private use. The remaining 16 codepoints are currently reserved, but in future have been proposed to be re-assigned to standard definitions.

The codepoints take their names from the Per Hop Behaviour (PHB) or forwarding treatment associated with them. PHBs include the default Best Effort (BE), AF1-4 (Assured Forwarding 1–4), which defines packet drop procedures and Expedited Forwarding (EF). Backwards compatibility with the earlier ToS specification was preserved by mapping eight codepoints (Class Selectors - CS0-7) to the old ToS precedence classes.

A network implementing DiffServ always assigns a better treatment to EF compared to AF or BE traffic, i.e. PHBs are ordered. This ordering presents a potential obstacle to inter-domain use if a DSCP modification were to remap the EF DSCP to a DSCP for a lower class, while leaving other DSCP values unchanged. This would cause the EF traffic to suffer priority inversion - a broken behaviour where a higher priority is remarked to a lower priority, while other priorities are left intact and not remarked. Consistent DSCP remapping needs to be designed to avoid this problem.

2.2. Characterisation of DSCP/ToS in internet

The literature on Internet traffic analysis is very rich. However, few studies have focused on DSCP/ToS measurements. Early analysis of ToS [8,9] evaluated the distribution of codepoints in the Internet, noting the scarce use of non-default traffic classes. However, these papers focused on ToS usage, rather than characterising pathologies that limit the usability of traffic classification.

More recent papers provide more insight on DSCP marking. A study [10] of 14,373 routers using Tracebox [11] reported a per-hop modification ratio of the DiffServ field of 5.75%. This sought to quantify middlebox interference, and hence did not investigate the detail of these modifications. A similar per-hop DSCP modification ratio has been reported [12] from analysing quotations resulting from ICMP probes to 84,393 web servers via tcptraceroute. This found an in-flight modification ratio of 2.9% for the ToS byte. A smaller study using Fling [13] reported high disruption to DSCP value in certain wireless access networks, observing connectivity failures, packet drops and DSCP modification dependent upon the DSCP. Unlike our study, these studies neither provided statistics on the affected DSCP/ToS fields nor did they attempt to classify the remarking pathologies.

Studies of ECN support [14–16] also help identify routers that continue to use ToS semantics. In 2015, results using PATHspider [15] showed 2.1% of IPv4 and 18.1% of IPv6 sampled hosts negotiated ECN, but never generated an ECT-marked packet. This suggests a potential bleaching of the ECN codepoint. This figure was later confirmed by a study for UDP traffic [17], which reported 2% of paths towards 2500 sampled servers were unreachable when the ECN field was non-zero.

2.3. Use of DiffServ by network operators

There are several approaches an operator may utilise the DSCP in packets forwarded through their domain:

- An operator could assign all traffic to single PHB, irrespective of the DSCP.
- An operator could forward the DSCP unchanged and map traffic to a set of PHBs (e.g., using MPLS).
- An operator could implement a policy that remaps specific DSCP values to a different DSCP value, and then assign that traffic to the corresponding PHB within their network.
- An operator could drop packets with a DSCP that is not supported in their network. However, this is inconsistent with Nichols et al. [1] that recommends domains forward unassigned DSCP values without change.

In the first two approaches the DSCP is carried transparently through the DiffServ domain. Our analysis in Section 4 shows that in the Internet backbone DSCP values are often transported transparently.

An operator that chooses to offer customers a restricted set of service classes (e.g., one supporting residential customers), may remap all traffic to a DSCP corresponding to the PHB for the service to which the customer subscribed. The same operator could support multiple service classes for enterprise customers, preserving the DSCP value as it cross their domain. Examples include the AT&T managed Internet service for business [18] and Comcast Xfinity,[1] which include voice services, Metro-Ethernet business services, IP cable and broadband Internet over a shared infrastructure, where the DSCP is used to prioritize traffic.

Our analysis is unable to distinguish between network operators that ignore a DSCP and operators that use this to assign traffic to a PHB. However, since both methods do not modify the DSCP value, they do not impede further use of the DSCP values at other points along the end-to-end path.

An operator that remaps the DSCP results in an observable pathology that can (if inconsistently applied) impact DSCP usability at other places on the path. While we did observe various examples of remarking, this practice was not as widespread as previously suggested (Section 4). We did not observe consistent remapping of traffic with a specific DSCP.

3. Methodology

Many tools are available for analysing Internet traffic. Our tests needed a tool able to efficiently process a large number of DSCP probes. This led us to design PathTrace, a tool to explore DSCP modification pathologies.

PathTrace exploits a mechanism similar to traceroute to infer

[1] https://www.xfinity.com/policies.

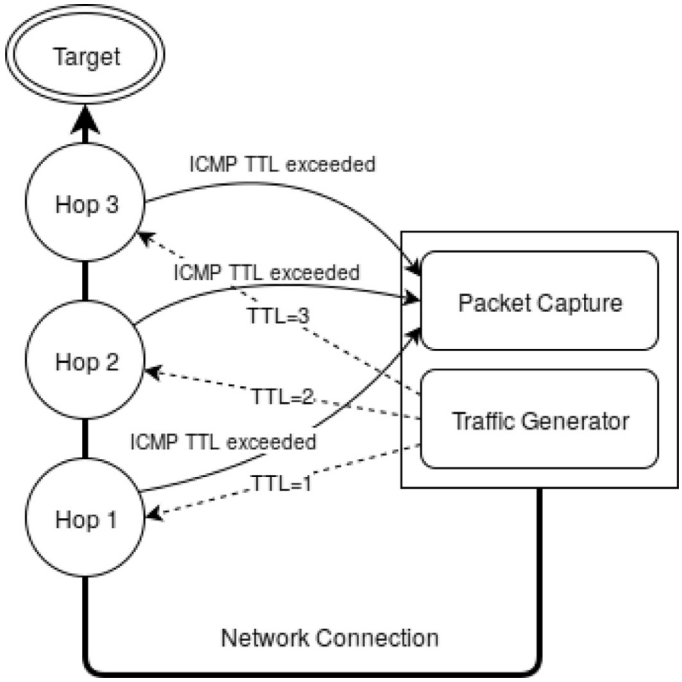

Fig. 1. The PathTrace tool.

information from quotations returned in Internet Control Message Protocol (ICMP) packets. All IP routers are required to decrement the TTL (IPv6 hop count) field and discard a packet if the TTL reaches zero. A router ought to also send an ICMP type 11 (respectively an ICMPv6 type 3 in IPv6) to notify the source of the discarded packet. The ICMP message includes a quotation of the IP header of the discarded packet. PathTrace detects modifications to DSCP value by comparing the quotations in two ICMP messages with consecutive TTL values.

Fig. 1 illustrates the architecture of PathTrace. The Traffic Generator is built upon the packet "forging and dissection" library Scapy.[2] Returning ICMP quotations are first stored in a file by the Packet Capture component, then analysed after the end of probing.

Other tools (e.g., [11,19]) have similarly extended traceroute to analyse ICMP quotations. In particular, Tracebox detects middleboxes [11] by analysing packet modifications including changes in the DSCP value. Fling is a tool that measures packet manipulation on the path to a destination server [13]. Other active measurement platforms, such as RIPE Atlas [20], OONI [21], or Netalyzr [22], also measure performance and/or connectivity between a pair of endpoints.

However, the existing tools could not be systematically used to test a range of DSCP values without adding significant functionality and optimising to reduce processing overhead. While similar to Tracebox, our tool is fast and lightweight. It stores probes and captured packets for post-processing. Unlike Fling and OONI, our tool does not require a central server and is not dependent on crowd-sourced measurements.

3.1. Experiment design and limitations

A known issue in using traceroute results when encountering per-packet (or per-flow) load-balancing routers. These can forward probes along different paths, for instance using equal cost multiple path routing, confusing processing of ICMP responses [19]. In our tests, packets probing different DSCP values were sent with the same 5-tuple to ensure that probes were not affected by this form of load balancing.

Another complication arises when traceroute is used over tunnels and with hidden nodes that also do not decrement the TTL [23,24]. MPLS tunnels in particular are common and unless the MPLS routers are instructed to reveal the internal structure of the tunnel (e.g using the `ttl-propagate` option in Cisco routers [23]) the nodes within the tunnel remain invisible. To mitigate this problem, we considered only the DSCP modifications observed at adjacent hops, as opposed to those observed in isolation [12].

We space probes for the first 4 hops, to reduce the likelihood of being impacted by routers within the local network that limit the rate of ICMP packets

To minimise disruption, we used well-known low ports normally allowed through firewalls. For the dataset "Mobile Edge", however, we used high numbered TCP and UDP ports, because low-numbered ports from the vantage point were firewalled and port 80 was redirected to a proxy service [25].

Traceroute suffers from the limitation that only routers closer to the vantage point can be probed accurately. As probes progress along the path, the DSCP may be modified reducing the ability to explore the full set of codepoints further along the path [26,27]. To evaluate how the pathologies are affected by our choice of vantage point, we present statistics for a vantage point in addition to aggregate statistics. This confirms common pathologies when probing routers in distant disjoint regions.

Finally, we note the need to consider interface aliases (when the same router is identified by more than one IP address), which could contribute bias in our results.

Our measurement campaigns explored use of UDP and TCP traffic. Web servers were used as targets, taking advantage of the existing comprehensive lists of popular web servers. DNS resolution was performed from a different location to all vantage points to avoid using pre-assigned mapping, and/or content caching and distribution mechanisms for popular servers (CDNs). Fetching web content is a common service, and it is possible that operators could have specific DiffServ policies in place. However, our results did not find any bias between the treatment of web and DNS traffic.

3.2. DSCP datasets

To cross-validate our findings, we considered three data sets. The list of target destinations was drawn from the top 100,000 websites

[2] http://www.secdev.org/projects/scapy/.

from the Cisco Umbrella 1 million[3] and Alexa Top 500 Sites. Webserver names were resolved using Hellfire.[4] When a name resolved to more than one address, one of the addresses was selected at random, except for the final campaign of the dataset "PathTrace" where one IPv4 and one IPv6 addresses were selected to diversify the sampling.

Dataset A ("PATHspider") was built between January and July 2017 extending to all non-zero codepoints the PATHspider measurements in Learmonth et al. [28] probing a random selection of 100,000 websites. This dataset includes additional measurements using DSCP 2 and DSCP 46 from eight Digital Ocean datacentres.[5]

Dataset B ("PathTrace") consisted of three measurement campaigns. The first campaign (*dataset B.1*) ran in December 2016 from three Digital Ocean data centres towards a list of 100 targets for a total of 300 source-destination pairs. The second campaign (*dataset B.2*) in February 2017 targeted 200 web servers using TCP port 80 from eight locations counting 1440 source-destination pairs. The third campaign (*dataset B.3*) ran in July 2017 towards the top 500 web servers in the Cisco top website list, using one IPv4 and IPv6 address for each target, for a total of 961 unique IP addresses. The ECN field of probe packets was set to 11 to infer correlations between ECN and DSCP modification.

Dataset C ("Mobile Edge") used PathTrace and featured data collected for 100 Webserver from the Alexa top websites list from 107 mobile vantage points within the European MONROE platform [29], for a total of 9202 different source-destination pairs. A range of DSCP values were sent using both UDP and TCP. The measurements were completed between September 2016 and January 2017.

3.3. DSCP pathologies

Our analysis revealed a number of recurrent DSCP modification pathologies, described below. The following section provides a quantitative analysis of the various pathologies.

- *DSCP bleaching*: This pathology resets the DSCP field to zero. This remaps all flows to the default traffic class.
- *ToS bleaching*: This pathology resets to zero the upper 3 bits of the DSCP field (the former ToS precedence field), leaving other bits unchanged. This behaviour is distinctive of non-DiffServ aware routers.
- *ToS bleaching except CS6/CS7*: This pathology is a variant of ToS bleaching, where the ToS precedence field is reset only if the ToS is not 110 or 111. These markings correspond to the CS6 (Network Control) and CS7 (Internetwork Control) codepoints, which identify critical Internet traffic. This also is an indication of a non-DiffServ aware router.
- *DSCP remarking and multiple remarking*: This pathology resets the DSCP field to a specific codepoint or to a pool of few codepoints. This may be a result of DiffServ traffic conditioning.

4. Results

4.1. DSCP remarking pathologies

To characterise DSCP modification pathologies, we considered the remarking behaviour of 3040 IPv4 and 1093 IPv6 routers in dataset B.2 and B.3. A router was considered only when it could be probed by at least 30 distinct codepoints. This allowed us to evaluate the router behaviour across a sufficiently large range of DSCP values to clearly identify the remarking pathology. Each probes sent TCP packets encapsulated in IPv4 or IPv6 datagrams. Table 1 reports he number of routers encountered in each dataset for each pathology, as well as the 95% confidence interval in brackets. Table 2 presents the percentage of routers with pathologies using each vantage point.

Table 1
DSCP modification pathologies.

Dataset B.2 *routers*=918		Dataset B.3 *routers* = 2122		Description IPv4
hosts	C.I. (%)	hosts	C.I. (%)	
694	(73–78)	1555	(71–75)	Transparent
10	(0.4–1.9)	81	(3.0–4.7)	Reset DSCP
102	(9.2–13)	156	(6.3–8.5)	Reset ToS prec.
33	(2.4–4.8)	54	(1.9–3.3)	Reset ToS prec. CS6/CS7
15	(0.9–2.5)	4	(0.0–0.4)	ToS prec. remap
10	(0.4–1.9)	59	(2.1–3.5)	DSCP remark
7	(0.2–1.4)	54	(1.9–3.3)	DSCP multiple remark
47	(3.7–6.5)	159	(6.4–8.6)	Other remarking
routers = 93		*routers* = 1000		IPv6
hosts	C.I. (%)	hosts	C.I. (%)	
85	(85–97)	886	(87–91)	Transparent
0	(0–3.2)	6	(0.2–1.1)	Reset DSCP
0	(0–3.2)	5	(0.1–0.1)	Reset ToS prec.
0	(0–3.2)	3	(0–0.7)	Reset ToS prec. CS6/CS7
2	(0–5.3)	30	(2–4.1)	DSCP remark
0	(0–3.2)	36	(2.5–4.8)	DSCP multiple remark
6	(2.1–12)	34	(2.3–4.6)	Other remarking

Table 2
Percentage for IPv4 DSCP modification pathologies per vantage point, dataset B.2.

AMS	BLR	FRA	LON	NYC	SFO	TOR	%
65.3	80	66.3	83.3	77.4	92.1	67.6	Transparent
1.8	5.4	2.6	6.2	6.3	2.5	7	Reset DSCP
24.6	2.6	16.1	3.6	10.1	0.7	15.4	Reset ToS prec.
3.75	3.8	4.8	0.8	2.7	0.3	1.4	Reset ToS prec. CS6/CS7
1.8	3.0	6.1	3.1	0.4	2.2	2.8	DSCP remark
2.1	0.4	3.4	2.2	2.7	1.4	5.6	DSCP multiple remark
320	420	229	354	444	827	71	Total routers

Almost one-fourth of IPv4 routers exhibited some pathological behaviour in DSCP remarking, with a significant number of routers implementing ToS bleaching (around 11% and 7% in dataset B.2 and B.3 respectively). This is clear evidence of routers continue to be configured with policies based on ToS semantics.

A router configured to use ToS semantics can utilise the ToS Precedence field to categorise traffic. A policy to disable such class-based flow management, could reset the 3 highest bits of the ToS/DSCP without updating the remainder of the field. This is however problematic when routers use DiffServ semantics, because it can result in unrecognized DSCP values for the remainder of the path and can result in priority inversion. This pathology was encountered from all vantage points, with a distribution that varied between 0.7% and 25%.

ToS bleaching (except for C6 and C7) was observed from all vantage points for around 3% of routers. This varied between 0.4% and 4.8% depending on vantage point and is also an indication of old or misconfigured routers. All vantage-point dependent variations can be attributed to using geographically diverse locations, connected via different providers.

A small proportion of IPv4 routers (around 1% and 4% respectively in dataset B.2 and B.3) implemented DSCP bleaching. We encountered this pathology for all vantage points tested, with between 1.8% and 7% variation depending on the vantage point. Resetting the DSCP may result from traffic conditioning at the edge of a DiffServ domain, or where no other policy is in place between two operators [30]. We observe this only in a small percentage of routers. This is less prevalent that suggested in [2]. Since DSCP-bleaching does not affect DSCP semantics (e.g., does not cause priority inversion), the impact of this is less disruptive than that caused by ToS precedence bleaching.

A number of other pathologies were observed. These include DSCP

[3] https://umbrella.cisco.com/blog/2016/12/14/cisco-umbrella-1-million/.
[4] https://github.com/irl/hellfire.
[5] https://www.digitalocean.com/.

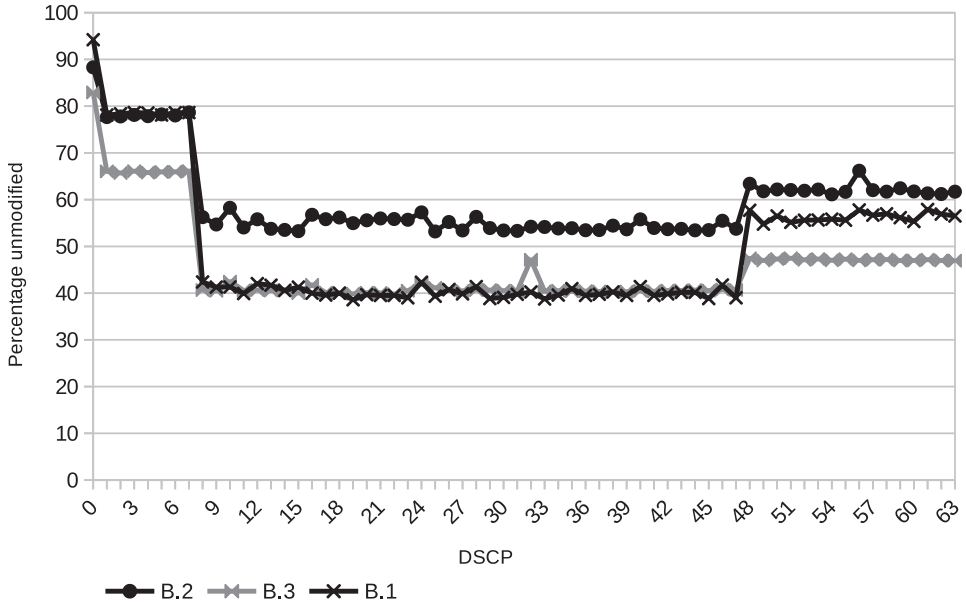

Fig. 2. Percentage of paths with no DSCP modification.

remarking and multiple remarking, mapping the ToS field to a different ToS class (*ToS remapping*), and other remarking pathologies involving only a range of DSCP values. Together these cases represent between 8.8% and 11% of tested routers. As anticipated, multiple remarking appears to result at the peering between DiffServ domains.

Remarking to a single or multiple codepoints was observed from all vantage points, varying between around 0.5% and 6% depending on vantage point. ToS remapping was observed in routers from one vantage point only, in campaign B.3.

Table 1 reports statistics for IPv6 routers. Although based on a smaller number of routers, these figures clearly show a much larger proportion of IPv6 routers transparently propagate the DSCP (around 90%). However, some remarking (< 1%) to zero, to a single codepoint (around 3%) or multiple codepoints(around 3%) was observed from all vantage points, with a vantage point dependent variation between 0.6% and 10% for single remarking and 1% and 6% for multiple remarking. Perhaps surprisingly, a small number of IPv6 routers (< 1%) were seen to employ ToS semantics, albeit observed only by three vantage points.

4.2. Preservation of the DSCP across networks

The remarking behaviour of routers can be detected by observing the DSCP value at the end of a path. Fig. 2 displays the percentage of transparent paths as a function of the original codepoint, comparing measurements in each of the datasets B. To avoid local bias, this only considers paths longer than three hops.

Excluding the default DSCP, which is delivered unchanged over more than 80% of paths, the diagram displays three plateaus: The plateau between DSCP 1 and 8 represents codepoints starting with '000'. These codepoints are unaffected by ToS precedence bleaching and therefore have a higher chance of not being changed by the path. Similarly, the rightmost plateau between DSCP 48 and 63 represents codepoints starting with '110' or '111'. These codepoints survive a ToS precedence bleaching that preserves CS6 and CS7. The remaining codepoints have a lower probability of survival and are unchanged over less than 40% of the paths.

Table 3
Percentage of remarking at the last hop of the mobile network (dataset C).

init. DSCP	BE	*Unch.*	DSCP 6	AF11	*Others*
BE	(73)	73	8.9	10	7.6
DSCP 3	60	8	12	11	8.1
CS1	54	36	2.4	2.1	5.7
AF11	54	38	2.4	(38)	5.9
EF	48	36	2.4	2.1	12

Table 4
Number of connections and percentage success/failure measured by PATHspider.

	# Conn. attempts	%
Both succeeded	12M	99.99
Both failed	5430	0.01
Baseline succeeded	3430	0.01
Test succeeded	6430	0.01

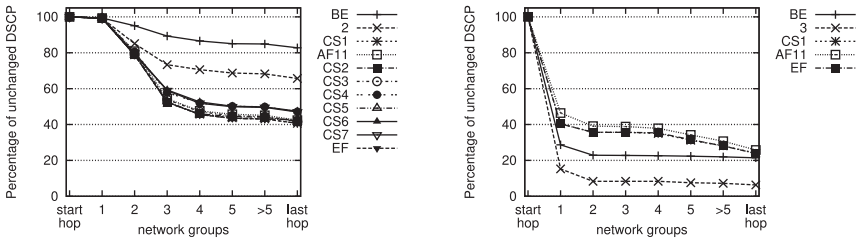

Fig. 3. Percentage of unchanged DSCP value for networks traversed.

Table 5
Remarking for each DSCP modification pathology using TCP and UDP.

TCP *routers* = 449 C.I. (%)		UDP *routers* = 449 C. I. (%)		Description Dataset B.1
335	(71–78)	337	(71–79)	Transparent
16	(2.0–5.3)	16	(2.0–5.3)	Reset DSCP
55	(9.4–15)	53	(8.9–15)	Reset ToS prec.
31	(4.7–9.3)	30	(4.5–9.1)	Reset ToS prec. CS6/CS7
12	(1.2–4.2)	13	(1.6–4.6)	Other remarking
routers = 507 C. I. (%)		*routers* = 393 C. I. (%)		Dataset C
418	(79–86)	325	(79–86)	Transparent
26	(3.4–7.1)	22	(3.6–7.9)	Reset DSCP
24	(3.0–6.7)	18	(2.5–6.9)	Reset ToS prec.
39	(5.5–10)	28	(4.6–9.7)	Other remarking

Fig. 3 illustrates the observed DSCP modifications along the path. The figure shows the percentage of DSCP values that remain unchanged, as a function of the number of *network groups* traversed. A network group is as a set of contiguous hops sharing a /16 IPv4 address prefix. We chose this grouping because we modifications tended to result at the peering points between domains rather than within domains. Fig. 3(a) refers to dataset B.3 featuring vantage points in Digital Ocean networks, whereas Fig. 3(b) refers to measurements in dataset C from mobile edge networks.

Nearly all the DSCP values traverse the first network without modification (Fig. 3(a)). This suggests our measurements are not biased by the infrastructure of our chosen vantage points. Most remarking occurs at the second and third network peering within backbone networks or at an Internet Exchange Point. Network equipment at Internet exchanges route packet across ASNs and can therefore appear in various network address spaces. For this reason, it is not always possible to identify which AS caused a modification.

4.3. DSCP at the mobile edge

This subsection focuses on mobile network vantage points. These measurements present a different set of pathologies and remarking behaviours compared to the core Internet. Fig. 3b shows that most DSCP values are remarked on entry to the mobile network. The percentage of unchanged codepoints falls below 40% immediately after the first network group. The DSCP values are coloured with specific non-zero codepoints, indicating an exclusive use of the DSCP within the mobile network. For example, many codepoints were remapped to the AF class (e.g., AF13 or AF12) irrespective of their initial value. This remarking seems to not be based on the DSCP value, with DSCP 0 often also remarked.

Table 6
DSCP remarking at the last hop when injecting EF probes using TCP and UDP.

TCP *paths* = 581 C. I. (%)		UDP *paths* = 581 C. I. (%)		DSCP at last hop Dataset B.1
223	(34–42)	225	(35–42)	BE
281	(44–52)	278	(44–52)	EF
46	(5.9–10)	49	(6.2–10.8)	6
14	(1.2–3.8)	14	(1.2–3.8)	CS1
7	(0–2.2)	7	(0.3–2.2)	41
10	(0–2.9)	8	(0.5–2.4)	Others
paths = 291 C. I. (%)		*paths* = 291 C. I. (%)		Dataset C
38	(9.3–17)	39	(9.6–18)	BE
148	(45–57)	147	(45–56)	EF
92	(26–37)	94	(27–38)	6
5	(0.3–3.4)	5	(0.3–3.4)	CS1
3	(0–2.4)	3	(0–2.4)	AF21
3	(0–2.4)	3	(0–2.4)	Others

We observed a second remarking at the edge of the network, before packets leave the mobile domain. Each column in Table 3 shows the percentage of each remarked DSCP value on leaving the mobile network compared to the DSCP value at entry. In most cases, DSCP values were bleached (i.e. reset to zero) when packets left the mobile domain.

The two-stage remarking behaviour, while typical of a mobile network, was not found in any of our core measurement datasets. It resulted in very few DSCP values (less than 5%) arriving unchanged through the mobile network.

4.4. DSCP-related connectivity impairments

Dataset A was used to investigate potential end-to-end connectivity problems resulting from use of a specific DSCP. The test opened a connection using the default DSCP value (baseline case), followed by opening connections to the same target using each of the other 63 codepoints (test case). The test case failed if a connection could not be opened for at least one of the DSCP values. Table 4 reports the number of connections where only the baseline case succeeded, only the test case succeeded, and both succeeded or failed.

These results confirm that the ability of endpoints to connect to an endpoint is not affected by the DSCP value. The small amount of breakage (less than 0.01%) is attributed to other network failures, such as busy server rejections or momentary link failures. These results differ from ones reported in [13] where some breakage was observed for specific networks. In our case, tests were performed from a vantage in a network known to propagate DSCP transparently (the academic Janet network). Repeating the test from a further eight locations for codepoints 2 and EF in Digital Ocean also obtained similar results.

4.5. Transport-dependent changes to the DSCP

The datasets B.1 and C were used to determine whether the choice of transport protocol (TCP or UDP) affected the DSCP remarking along a path.

Table 5 reports a breakdown of the remarking ratio for each DSCP modification pathology for the set of routers in the dataset that were reached by at least eight distinct DSCP probes. The first and second column in Table 5 report the number of routers that introduced a certain pathology for TCP and UDP respectively along with the corresponding range of modification ratios (in percentage) between brackets.

The similarity of success for TCP and UDP is evident (a two-sample *t*-test produces *p*-values larger than 0.8). This demonstrates that current remarking is unrelated to the choice of transport protocol. The small discrepancy is attributed to other failures, such as temporary link failures or congestion loss of a probe packet.

A similar result was also found when evaluating DSCP remarking at the last hop. The first and second column in Table 6 report the number of DSCP values observed at the last hop when a probe was sent with an EF codepoint. Again, the small discrepancy is not attributed to the choice of transport protocol.

4.6. DSCP observed at the last hop

Fig. 4 presents our data in the form of a heat map. This plots the original DSCP value (vertical axis) against the DSCP value at the end of the observed path (horizontal axis). The strong diagonal line in the plot corresponds to DSCP values that remain unchanged across the path.

The vertical lines in the map indicate remarking to the DSCP value shown on the horizontal axis. Vertical lines can be identified for DF/BE (very strong), CS1 (very faint), AF11, AF21, CS3 and CS4. A significant number of routers remap all incoming codepoints to default (0). There is some remapping to other well-known codepoints.

One other diagonal pattern can be distinguished, spanning DSCP values 0–7. These repeat seven times in a clear pattern, consistent with

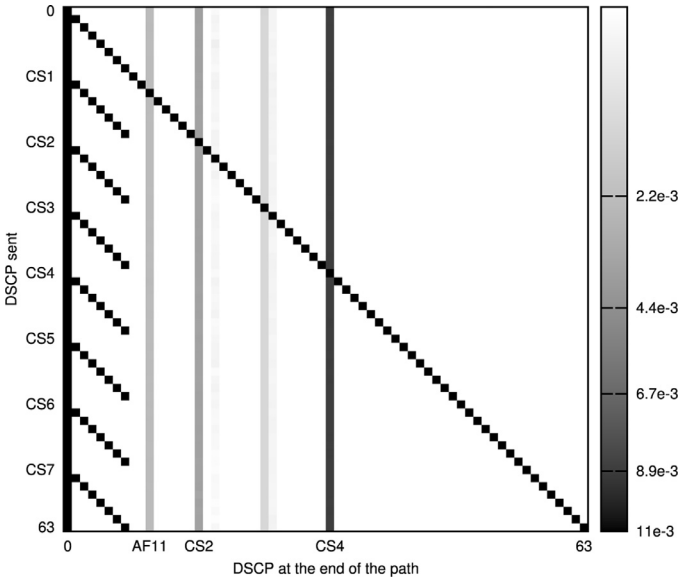

Fig. 4. Observed DSCP at the last hop, (n = 1740).

Table 7
Comparison of DSCP and ECN modifications in 3,086,977 adjacent hops.

	DSCP changed		DSCP unchanged	
	hops	(%)	hops	(%)
ECN cleared	2555	40%	6032	60%
ECN unchanged	255,421	8.2%	2,822,969	91.8%

bleaching of the three high-order bits (ToS bleaching).

4.7. Changes to the ECN field and DSCP

Dataset B.3 was used to evaluate the dependency between DSCP and ECN modification. The experiment sent probes with the ECN field set to 11 (congestion experienced - CE mark) and counted the number of hops where a transition to 00 (non-ECN mark) was observed, both with and without DSCP modification. Table 7 presents a 2 × 2 contingency table reporting the number of hops in each of the four cases (with DSCP or ECN modified or not). The number in brackets reports the percentage of hops with respect to the total for each row.

Although the number of router hops that clear the ECN field is small (0.3%), this data shows a clear dependency between clearing the field and changing the DSCP value (χ^2-test provides a p-value much larger than one). 8.2% of hops register a change of DSCP with no change to the ECN field. This proportion rises to 40% when the ECN field is cleared.

An analysis of the type of changes that occur when the ECN field is cleared reveals that the DSCP was reset in 52% of cases and remarked to a different DSCP in 48%. These results confirm the findings in [31] that ECN clearing is related to modification of the entire ToS byte. Although more disruptive to DiffServ, resetting the ToS precedence is safer for ECN because this does not alter the ECN field.

5. Discussion

99.99% of the tested paths offered connectivity irrespective of the DSCP value that was used for the packets. This result is encouraging for DiffServ deployment since it shows that DiffServ-based packet dropping within the core and server side networks is not common.

In considering whether DiffServ offers a useful QoS function, we examine how far the original QoS requirement is reflected by the DSCP as packets are forwarded. Our experiments observed that over 70% of routers pass the DSCP without modification. In the remaining cases, the DSCP is remarked. The discussion now explores several of these cases.

5.1. Impact of legacy ToS treatment

DSCP modification by routers using ToS semantics are the biggest barrier to survivability of commonly used DSCP values, but we found no evidence this results in loss. In the set of routers that changed the DSCP, the most prevalent pathology was to reset to zero the highest three bits of the DiffServ field. This pathology is attributed to IPv4 routers that implement obsoleted ToS Precedence bleaching. This DSCP manipulation pathology reduces the likelihood that packets receive the desired PHB for the remainder of their path. A small number of IPv6 routers exhibit a similar pathology.

There is a significant opportunity to improve end-to-end transparency by by updating these router configurations to use DiffServ semantics. We also expect the prevalence of this pathology to diminish as old equipment reaches the end of life.

5.2. DiffServ interconnection (Intercon)

Differentiated Services (DiffServ) InterconY.1566 defines a set of four common QoS classes and four auxiliary classes, to which DiffServ traffic may be mapped [2]. This targets operations between separately administered networks interconnected using the MPLS Short-Pipe tunnel mode and has the potential to extend consistent DiffServ treatment between DiffServ domains. The codepoints chosen do not set the highest two bits of the DSCP; with the exception of the default DSCP, but we note that this set of codepoints are in the range shown in Fig. 2 have the highest probability of DSCP modification as a packet traverses the path.

Intercon also recommends remarking unknown or unexpected codepoints to default (DSCP 0), assuming this practice is already widely deployed. However, our results for core and server paths did not show this was common practice for the network paths that we tested. Only a small number of routers remarked a subset of codepoints to DSCP 0 and remapped others differently. These routers may be closer to the edge, and therefore the number of codepoints seen was too few to be selected for analysis. Our data also only examined the core and data centre portions of the network path, and it may be that access providers have adopted other practices, about which we are unable to currently comment.

5.3. Comparison to mobile results

A different set of pathologies arise for mobile networks, which are much less transparent to the DSCP value. Inside the mobile networks that we studied, we observed remarking to several codepoints dependent on the country and mobile operator. The remarking was irrespective of the original DSCP, implying a remarking by a local policy.

The GSM Association guidelines [32] for interconnection of mobile backbones could in future help coordinate inter-domain use of DSCPs within mobile networks. However, we did not see evidence of these guidelines currently being implemented.

When packets leave the mobile domain, they are subject to the same pathologies as in the Internet core. The most prevalent (58% of packets) is to reset the DSCP to 0, before they traverse the remainder of the path.

5.4. Selecting a DSCP for applications

Results using PATHspider show that it is safe to enable DSCP marking for applications. There is very little evidence of packet loss due to using a specific codepoint. An application can expect to gain benefits from DiffServ locally, but current data suggests it is likely to experience remarking after a few hops. Within the core, routers using ToS semantics can also still lead to unrecognized codepoints that prevent packets from receiving the desired PHB in the later part of their path.

Of the standardised DSCP values, DSCP 0–7 were observed to be the least unchanged on core paths, which we suggest is due to the higher order 3 bits of the DSCP being already zero. We did not see significant evidence that using any other well-known codepoint will significantly increase/decrease the probability of successful DSCP end-to-end traversal, but note the remarking recommendations in Intercon, as a sign that unsupported DSCP marks could in future be remarked as default (DSCP 0).

For mobile networks, applications can expect to sometimes exploit DiffServ locally, potentially gaining benefit within the mobile network, but at the current time, we would expect remarking of most DSCP values after traversing a mobile network.

5.5. A DiffServ codepoint for scavenger traffic

Previous work suggested the use of a DSCP to identify traffic desiring a Lower Effort LE treatment (also known as scavenger class). RFC3662 [33] suggests using CS1 (DSCP 8) for this traffic, a marking that has been used in Internet2 [34]. However, while this code point is permitted in DiffServ, it is at odds with the normal priority of CS class markings, and has not been officially assigned by IANA for this purpose.

We therefore explore the suitability of using DSCP 8, based on our understanding of DSCP modification pathologies in Internet. The results in Fig. 4 showing ToS precedence bleaching, indicate this codepoint has a 36.82% probability of being reset to DSCP 0, causing traffic to be treated with the PHB for the default class. This is arguably better than priority inversion, but does not realise the desired an LE treatment. It is therefore important that any use of DSCP 8 for background traffic does not rely solely on the DSCP for controlling the capacity used by the scavenger application.

Recent work at the IETF is revisiting the use of DSCP 8 and one proposal is use DSCP 2 for LE traffic. This codepoint has one of highest retention rates, we therefore expect it to be forwarded by routers without change. This DSCP is not subject to priority inversion. However, we expect other traffic (AF11, AF21, AF31, AF41) to be remarked by ToS precedence bleaching resulting in this codepoint, this would cause priority inversion for AF traffic and the possibility that this traffic may erroneously also be assigned to the LE PHB.

Although prevalence of ToS precedence bleaching is expected to diminish with time, priority inversion is nevertheless a serious concern. After understanding these concerns, the IETF is instead considering reassigning use of DSCP 1 as a marking for LE traffic.

5.6. Recommendations for DSCP usage in webRTC

WebRTC provides browsers and mobile applications with Real-Time Communications (RTC) capabilities via simple APIs. IETF work in support of WebRTC [35] recommends a set of DSCP values for general Internet use. This subsection briefly examines the pathologies for this set of codepoints. WebRTC is typically used as a peer-to-peer application and would therefore benefit from edge-to-edge support for DSCP markings.

The specification recommends using the default class (DSCP 0) for low priority, the EF class for voice, and a set of AF class markings for video traffic. Our results show that traffic with these markings was passed through the networks that we tested. Remarking to DSCP 0 and ToS bleaching could impact the ability of the remote endpoint to observe the desired DSCP, in both core and mobile networks.

The proposed specification currently recommends use of CS1 for traffic with a “very low” application priority. A future standardised LE codepoint may be more suitable for this traffic.

6. Future work

Results show transparency with respect to packet traversal, but still display unwanted pathologies as the DiffServ field is changed on an Internet path. The measurement technique presented in this paper may help identify legacy routers that need to be replaced or reconfigured to avoid these undesirable pathologies. The current prevalence of such router configurations suggests that these measurements could also usefully be repeated in future years to track whether this problem reduces as predicted.

Our measurement results for the core should encourage increased attention to enabling DiffServ in the access part of the Internet path. Our exploration of DSCP modification pathologies at the edge of the Internet was limited to mobile networks. We therefore encourage experimentation to understand DSCP remarking pathologies across a range of access equipment networks.

We encourage operators to continue to deploy PHBs to which DSCP packet markings can be mapped, and to make this information available. However, we were unable to test whether PHBs had been deployed in the networks we tested, nor could we comment on efforts by operators to implement conditioning at the boundaries between DiffServ domains. Measurements examining the forwarding treatment received by packets are by their nature more disruptive than the tests described in this paper, and may be hard to verify without congestion information at the time of measurement.

7. Conclusion and next steps

This paper presents a new tool for observing DSCP modification pathologies and provides new large-scale measurements using fixed-core and mobile edge networks. Our results examine a range of DSCP values and modification pathologies as packets traverse an end-to-end path. While we observed few cases where networks discard packets with a specific codepoint, the more significant result is that many networks do modify the DSCP value. Even so, we recommend applications to set a DSCP and provide specific recommendations.

While there is evidence of operator configuration using DiffServ, much of the observed remarking appears to arise from routers configured to use historic ToS semantics. In some cases, this results in priority inversion. The strong recommendation is to reconfigure and/or upgrade these routers, to provide greater opportunity for using DiffServ across an entire network path. We also recommend continued measurement of DSCP remarking both in the core/server portions of the network and to characterise access networks.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.comcom.2018.05.016.

References

[1] K. Nichols, S. Blake, F. Baker, D. Black, Definition of the differentiated services field (DS Field) in the IPv4 and IPv6 headers, 1998, (RFC 2474).
[2] R. Geib, D. Black, Diffserv-interconnection classes and practice, 2017, (RFC 8100 (Informational)).
[3] C. Filsfils, J. Evans, Deploying diffserv in backbone networks for tight sla control, IEEE Internet Comput. 9 (1) (2005) 66–74, http://dx.doi.org/10.1109/MIC.2005.12.
[4] R. Bless, A lower effort per-hop behavior (LE PHB), Work in progress, Internet Engineering Task Force, 2017.
[5] A. Custura, G. Fairhurst, A. Venné, Exploring dscp modification pathologies in mobile edge networks, Proc. of IEEE/IFIP Workshop on Mobile Network Measurement, (2017).
[6] K. Ramakrishnan, S. Floyd, D. Black, The addition of explicit congestion notification (ECN) to IP, 2001, RFC 3168 (IETF).
[7] 2017.
[8] F. Li, N. Seddigh, B. Nandy, D. Matute, An empirical study of today's internet traffic for differentiated services IP QoS, in: IEEE ISCC 2000, Washington (USA), pp. 207–213.
[9] D. Murray, T. Koziniek, The state of enterprise network traffic in 2012, in: IEEE APCC 2012, Berlin, pp. 179–184.
[10] K. Edeline, B. Donnet, Towards a middlebox policy taxonomy: Path impairments, IEEE INFOCOM'15, Hong Kong, (2015), pp. 402–407.
[11] G. Detal, B. Hesmans, O. Bonaventure, Y. Vanaubel, B. Donnet, Revealing middlebox interference with tracebox, (2013), http://dx.doi.org/10.1145/2504730.2504757.
[12] D. Malone, M. Luckie, Analysis of ICMP Quotations, Proc. of PAM'07, Berlin, (2007), pp. 228–232.
[13] R. Barik, M. Welzl, A. Elmokashfi, How to say that you're special: Can we use bits in the IPv4 header? Proceedings of ACM ANRW'16, Berlin, (2016), pp. 68–70.
[14] M. Kühlewind, S. Neuner, B. Trammell, On the state of ECN and TCP options on the internet, Proceedings of PAM'13, Hong Kong, (2013), pp. 135–144.
[15] B. Trammell, et al., Enabling internet-wide deployment of explicit congestion notification, Proceedings of PAM'15, Brooklyn (USA), (2015), pp. 193–205.
[16] P. Richter, et al., Distilling the internet's application mix from packet-sampled traffic, Proc. of PAM'15, New York (USA), (2015), pp. 179–192.
[17] S. McQuistin, C.S. Perkins, Is explicit congestion notification usable with UDP? in: Proc. of IMC'15, Tokyo, pp. 63–69.
[18] AT&T, Class of service data collection document for at&t managed internet service (mis), 2010.
[19] B. Augustin, et al., Avoiding traceroute anomalies with paris traceroute, in: Proceedings of IMC '06, Rio de Janeriro (Brazil), pp. 153–158.
[20] R.N. Staff, RIPE atlas: a global internet measurement network, Internet Protocol J. 18 (3) (2015).
[21] A. Filasto, J. Appelbaum, OONI: open observatory of network interference. Proc. of FOCI (Usenix), Bellevue (USA), (2012).
[22] C. Kreibich, N. Weaver, B. Nechaev, V. Paxson, Netalyzr: illuminating the edge network, Proc. of ACM IMC'10, (2010), pp. 246–259.
[23] B. Donnet, M. Luckie, P. Mérindol, J. Pansiot, Revealing MPLS tunnels obscured from traceroute, Comput. Commun. Rev. 42 (2) (2012) 87–93.
[24] T. Flach, E. Katz-Bassett, R. Govindan, Quantifying violations of destination-based forwarding on the internet, in: Proc. of ACM IMC'12, Boston (USA), pp. 265–272.
[25] I.R. Learmonth, A. Lutu, G. Fairhurst, D. Ros, Ö. Alay, Path transparency measurements from the mobile edge with PATHspider, in: Proc. of IEEE/IFIP TMA'17, Dublin, pp. 1–6.
[26] D. Achlioptas, A. Clauset, D. Kempe, C. Moore, On the bias of traceroute sampling: or, power-law degree distributions in regular graphs, J. ACM (JACM) 56 (4) (2009) 21:1–21:28.
[27] M. Luckie, et al., Measured impact of crooked traceroute, Comput. Commun. Rev. 41 (1) (2011) 14–21.
[28] I.R. Learmonth, B. Trammell, M. Kühlewind, G. Fairhurst, PATHspider: A tool for active measurement of path transparency, Proc. of ACM ANRW'16, Berlin, (2016), pp. 62–64.
[29] MONROE - measuring mobile broadband networks in Europe.
[30] A. Terzis, B. Braden, S. Vincent, L. Zhang, RSVP diagnostic messages, 2000, (RFC 2745 (Proposed Standard)).
[31] A.M. Mandalari, A. Lutu, B. Briscoe, M. Bagnulo, O. Alay, Measuring ECN++: Good News for ++, Bad News for ECN over Mobile, IEEE Commun. Mag. 56 (3) (2018) 108–186.
[32] G. Association, Guidelines for ipx provider networks version 12.0(2016). http://www.gsma.com/newsroom/wp-content/uploads/IR.34-v12.0.pdf.
[33] R. Bless, K. Nichols, K. Wehrle, A lower effort per-domain behavior (PDB) for differentiated services, 2003, RFC 3662 (IETF).
[34] S. Shalunov, B. Teitelbaum, Qbone Scavenger Service (qbss) Definition, Internet2 Technical report, (2001). Proposed Service Definition.
[35] S. Dhesikan, D. Druta, P. Jones, C. Jennings, DSCP Packet Markings for WebRTC QoS, Work in progress, (2016). draft-ietf-tsvwg-rtcweb-qos.