TSVWG
Internet-Draft
Intended status: Informational
Expires: December 22, 2016

R. Geib, Ed.
Deutsche Telekom
D. Black
EMC Corporation

June 20, 2016

Diffserv-Interconnection classes and practice
draft-ietf-tsvwg-diffserv-intercon-06

Abstract

This document defines a limited common set of Diffserv PHBs and codepoints (DSCPs) to be applied at (inter)connections of two separately administered and operated networks, and explains how this approach can simplify network configuration and operation. Many network providers operate MPLS using Treatment Aggregates for traffic marked with different Diffserv PHBs, and use MPLS for interconnection with other networks. This document offers a simple interconnection approach that may simplify operation of Diffserv for network interconnection among providers that use MPLS and apply the Short-Pipe tunnel mode.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 22, 2016.

Table of Contents

1. Introduction

Diffserv has been deployed in many networks; it provides differentiated traffic forwarding based on the Diffserv Codepoint (DSCP) field [RFC2474]. This document defines a set of common Diffserv QoS classes (Per Hop Behaviors, PHBs)and code points at interconnection points to which and from which locally used classes and code points should be mapped.

As described by section 2.3.4.2 of RFC 2475, remarking of packets at domain boundaries is a Diffserv feature [RFC2475]. If traffic marked with unknown or unexpected DSCPs is received, RFC2474 recommends forwarding that traffic with default (best effort) treatment without changing the DSCP markings to better support incremental Diffserv deployment in existing networks as well as with routers that do not support Diffserv or are not configured to support it. Many networks

do not follow this recommendation, and instead remark unknown or unexpected DSCPs to zero upon receipt for default (best effort) forwarding in accordance with the guidance in RFC 2475 [RFC2475] to ensure that appropriate DSCPs are used within a Diffserv domain. This draft assumes that latter approach by defining additional DSCPs that are known and expected at network interconnection interfaces.

This document is motivated by requirements for IP network interconnection with Diffserv support among providers that operate MPLS in their backbones, but is applicable to other technologies. The operational simplifications and methods in this document help align IP Diffserv functionality with MPLS limitations resulting from the widely deployed Short Pipe tunnel model for operation [RFC3270]. Further, limiting Diffserv to a small number of Treatment Aggregates can enable network traffic to leave a network with the DSCP value with which it was received, even if a different DSCP is used within the network, thus providing an opportunity to extend consistent QoS treatment across network boundaries.

In isolation, use of a defined set of interconnection PHBs and DSCPs may appear to be additional effort for a network operator. The primary offsetting benefit is that mapping from or to the interconnection PHBs and DSCPs is specified once for all of the interconnections to other networks that can use this approach. Absent this approach, the PHBs and DSCPs have to be negotiated and configured independently for each network interconnection, which has poor administrative and operational scaling properties. Further, consistent end-to-end QoS treatment is more likely to result when an interconnection code point scheme is used because traffic is remarked to the same PHBs at all network interconnections.

The interconnection approach described in this document (referred to as Diffserv-Intercon) uses a set of PHBs and MPLS treatment aggregates along with a set of interconnection DSCPs allowing straightforward rewriting to domain-internal DSCPs and defined DSCP markings for traffic forwarded to interconnected domains. The solution described here can be used in other contexts benefitting from a defined interconnection QoS interface.

The basic idea is that traffic sent with a Diffserv-Interconnect PHB and DSCP is restored to that PHB and DSCP at each network interconnection, even though a different PHB and DSCP may be used by each network involved. The key requirement is that the network ingress interconnect DSCP be restored at network egress, and a key observation is that this is only feasible in general for a small number of DSCPs. Traffic sent with other DSCPs can be remarked to an interconnect DSCP or dealt with via additional agreement(s) among the operators of the interconnected networks; remarking in the absence of

additional agreement(s) when the MPLS Short Pipe model is used for reasons explained in this document.

In addition to the common interconnecting PHBs and DSCPs, interconnecting operators need to further agree on the tunneling technology used for interconnection (e.g., MPLS, if used) and control or mitigate the impacts of tunneling on reliability and MTU.

1.1. Related work

In addition to the activities that triggered this work, there are additional RFCs and Internet-drafts that may benefit from an interconnection PHB and DSCP scheme. RFC 5160 suggests Meta-QoS-Classes to enable deployment of standardized end to end QoS classes [RFC5160]. The Diffserv-Intercon class- and codepoint scheme is intended to complement that work (e.g. by enabling a defined set of end-to-end QoS service classes).

BGP signaling Class of Service at interconnection interfaces by BGP [I-D.knoll-idr-cos-interconnect], [ID.ietf-idr-sla] is complementary to Diffserv-Intercon. These two BGP documents focus on exchanging SLA and traffic conditioning parameters and assume that common PHBs identified by the signaled DSCPs have been established (e.g., via use of the Diffserv-Intercon DSCPs) prior to BGP signaling of QoS.

1.2. Applicability Statement

This document is applicable to use of Differentiated Services for interconnection traffic between networks, and in particular to interconnection of MPLS-based networks. This document is not intended for use within an individual network, where the approach specified in RFC 5127 [RFC5127] is among the possible alternatives; see Section 3 for further discussion.

The Diffserv-Intercon approach described in this document simplifies IP based interconnection to domains operating the MPLS Short Pipe model for IP traffic, both terminating within the domain and transiting onward to another domain. Transiting traffic is received and sent with the same PHB and DSCP. Terminating traffic maintains the PHB with which it was received, however the DSCP may change.

Diffserv-Intercon may also be applied to the Pipe tunneling model [RFC2983], [RFC3270], but is not applicable to the Uniform tunneling model [RFC2983], [RFC3270].

1.3. Document Organization

This document is organized as follows: section 2 reviews the MPLS Short Pipe tunnel model for Diffserv Tunnels [RFC3270], because effective support for that model is a crucial goal of Diffserv-Intercon. Section 3 provides background on RFC 5127's approach to traffic class aggregation within a Diffserv network domain and contrasts it with the Diffserv-Intercon approach. Section 4 introduces Diffserv-Interconnection Treatment Aggregates, along with the PHBs and DSCPs that they use, and explains how other PHBs (and associated DSCPs) may be mapped to these Treatment Aggregates. Section 4 also discusses treatment of non-tunneled and tunneled IP traffic and MPLS VPN QoS considerations and handling of high-priority Network Management traffic is described. Appendix A describes how the MPLS Short Pipe model (penultimate hop popping) impacts QoS and DSCP marking for IP interconnections.

2. MPLS and the Short Pipe tunnel model

The Pipe and Uniform models for Differentiated Services and Tunnels are defined in [RFC2983]. RFC3270 adds the Short Pipe model in order to support MPLS penultimate hop popping (PHP) of Labels, primarily for MPLS-based IP tunnels and VPNs. The Short Pipe model and PHP have subsequently become popular with network providers that operate MPLS networks and are now widely used to transport non-tunneled IP traffic, not just traffic encapsulated in IP tunnels and VPNs. This has important implications for Diffserv functionality in MPLS networks.

RFC 2474's recommendation to forward traffic with unrecognized DSCPs with Default (best effort) service without rewriting the DSCP has proven to be a poor operational practice. Network operation and management are simplified when there is a 1-1 match between the DSCP marked on the packet and the forwarding treatment (PHB) applied by network nodes. When this is done, CS0 (the all-zero DSCP) is the only DSCP used for Default forwarding of best effort traffic, and a common practice is to remark to CS0 any traffic received with unrecognized or unsupported DSCPs at network edges.

MPLS networks are more subtle in this regard, as it is possible to encode the provider's DSCP in the MPLS Traffic Class (TC) field and allow that to differ from the PHB indicated by the DSCP in the MPLS-encapsulated IP packet. If the MPLS label with the provider's TC field is present at all hops within the provider network, this approach would allow an unrecognized DSCP to be carried edge-to-edge over an MPLS network, because the effective DSCP used by the provider's MPLS network would be encoded in the MPLS label TC field

(and also carried edge-to-edge). Unfortunately this is only true for the Pipe tunnel model.

The Short Pipe tunnel model and PHP behave differently because PHP removes and discards the MPLS provider label carrying the provider's TC field before the traffic exits the provider's network. That discard occurs one hop upstream of the MPLS tunnel endpoint (which is usually at the network edge), resulting in no provider TC info being available at tunnel egress. To ensure consistent handling of traffic at the tunnel egress, the DSCP field in the MPLS-encapsulated IP header has to contain a DSCP that is valid for the provider's network, so that IP header cannot be used to carry a different DSCP edge-to-edge. See Appendix A for a more detailed discussion.

3. Relationship to RFC 5127

This document draws heavily upon RFC 5127's approach to aggregation of Diffserv traffic classes for use within a network, but there are important differences caused by characteristics of network interconnects that differ from links within a network.

3.1. RFC 5127 Background

Many providers operate MPLS-based backbones that employ backbone traffic engineering to ensure that if a major link, switch, or router fails, the result will be a routed network that continues to function. Based on that foundation, [RFC5127] introduced the concept of Diffserv Treatment Aggregates, which enable traffic marked with multiple DSCPs to be forwarded in a single MPLS Traffic Class (TC) based on robust provider backbone traffic engineering. This enables differentiated forwarding behaviors within a domain in a fashion that does not consume a large number of MPLS Traffic Classes.

RFC 5127 provides an example aggregation of Diffserv service classes into 4 Treatment Aggregates. A small number of aggregates are used because:

- The available coding space for carrying QoS information (e.g., Diffserv PHB) in MPLS (and Ethernet) is only 3 bits in size, and is intended for more than just QoS purposes (see e.g. [RFC5129]).

- The common interconnection DSCPs ought not to use all 8 possible values. This leaves space for future standards, for private bilateral agreements and for local use PHBs and DSCPs.

- Migrations from one Diffserv code point scheme to a different one is another possible application of otherwise unused QoS code points.

3.2. Differences from RFC 5127

Like RFC 5127, this document also uses four traffic aggregates, but differs from RFC 5127 in some important ways:

- It follows RFC 2475 in allowing the DSCPs used within a network to differ from those to exchange traffic with other networks (at network edges), but provides support to restore ingress DSCP values if one of the recommended interconnect DSCPs in this draft is used. This results in DSCP remarking at both network ingress and network egress, and this draft assumes that such remarking at network edges is possible for all interface types.

- Diffserv-Intercon suggests limiting the number of interconnection PHBs per Treatment Aggregate to the minimum required. As further discussed below, the number of PHBs per Treatment Aggreagate is no more than two. When two PHBs are specified for a Diffserv-Intercon treatment aggregate, the expectation is that the provider network supports DSCPs for both PHBs, but uses a single MPLS TC for the Treatment Aggregate that contains the two PHBs.

- Diffserv-Intercon suggests mapping other PHBs and DSCPs into the interconnection Treatment Aggregates as further discussed below.

- Diffserv-Intercon treats network control traffic as a special case. Within a provider's network, the CS6 DSCP is used for local network control traffic (routing protocols and OAM traffic that is essential to network operation administration, control and management) that may be destined for any node within the network. In contrast, network control traffic exchanged between networks (e.g., BGP) usually terminates at or close to a network edge, and is not forwarded through the network because it is not part of internal routing or OAM for the receiving network. In addition, such traffic is unlikely to be covered by standard interconnection agreements; rather, it is more likely to be specifically configured (e.g., most networks impose restrictions on use of BGP with other networks for obvious reasons). See Section 4.2 for further discussion.

- Because RFC 5127 used a Treatment Aggregate for network control traffic, Diffserv-Intercon can instead define a fourth traffic aggregate to be defined for use at network interconnections instead of the Network Control aggregate in RFC 5127. Network Control traffic may still be exchanged across network interconnections as further discussed in Section 4.2. Diffserv-Intercon uses this fourth traffic aggregate for VoIP traffic, where network-provided QoS is crucial, as even minor glitches are immediately apparent to the humans involved in the conversation.

4. The Diffserv-Intercon Interconnection Classes

At an interconnection, the networks involved need to agree on the PHBs used for interconnection and the specific DSCP for each PHB. This document defines a set of 4 interconnection Treatment Aggregates with well-defined DSCPs to be aggregated by them. A sending party remarks DSCPs from internal schemes to the interconnection code points. The receiving party remarks DSCPs to their internal scheme. The interconnect SLA defines the set of DSCPs and PHBs supported across the two interconnected domains and the treatment of PHBs and DSCPs that are not recognized by the receiving domain.

Similar approaches that use of a small number of traffic aggregates (including recognition of the importance of VoIP traffic) have been taken in related standards and recommendations from outside the IETF, e.g., Y.1566 [Y.1566], GSMA IR.34 [IR.34]and MEF23.1 [MEF23.1].

The list of the four Diffserv-Interconnect traffic aggregates follows, highlighting differences from RFC 5127 and suggesting mappings for all RFC 4594 traffic classes to Diffserv-Intercon Treatment Aggregates:

Telephony Service Treatment Aggregate: PHB EF, DSCP 101 110 and PHB VOICE-ADMIT, DSCP 101100, see [RFC3246],[RFC4594]and [RFC5865]. This Treatment Aggregate corresponds to RFC 5127s real time Treatment Aggregate definition regarding the queuing (both delay and jitter should be minimized), but this aggregate is restricted to transport Telephony Service Class traffic in the sense of RFC 4594 [RFC4594].

Bulk Real-Time Treatment Aggregate: This Treatment Aggregate is designed to transport PHB AF41, DSCP 100 010 (the other AF4 PHB group PHBs and DSCPs may be used for future extension of the set of DSCPs carried by this Treatment Aggregate). This Treatment Aggregate is intended for Diffserv-Intercon network interconnection of the portions of RFC 5127's Real Time Treatment Aggregate, that consume significant bandwidth. This traffic is expected to consist of the RFC4594 classes Broadcast Video, Real-Time Interactive and Multimedia Conferencing. This treatment aggregate should be configured with a rate queue (consistent with RFC 4594's recommendation for the transported traffic classes). By comparison to RFC 5127, the number of DSCPs has been reduced to one (initially). The AF42 and AF43 PHBs could be added if there is a need for three-color marked Multimedia.

Assured Elastic Treatment Aggregate This Treatment Aggregate consists of PHBs AF31 and AF32 (i.e., DSCPs 011 010 and 011

100). By comparison to RFC 5127, the number of DSCPs has been reduced to two. This document suggests to transport signaling marked by AF31 (e.g. as recommended by GSMA IR.34 [IR.34]). AF33 is reserved for extension of PHBs to be aggregated by this TA. For Diffserv-Intercon network interconnection, the following RFC 4594 service classes should be mapped to the Assured Elastic Treatment Aggregate: the Signaling Service Class (being marked for lowest loss probability), Multimedia Streaming Service Class, the Low-Latency Data Service Class and the High-Throughput Data Service Class.

Default / Elastic Treatment Aggregate: transports the default PHB, CS0 with DSCP 000 000. RFC 5127 example refers to this Treatment Aggregate as Aggregate Elastic. An important difference from RFC 5127 is that any traffic with unrecognized or unsupported DSCPs may be remarked to this DSCP. For Diffserv-Intercon network interconnection, the RFC 4594 standard service class and Low-priority Data service class should be mapped to this Treatment Aggregate. This document does not specify an interconnection class for RFC 4594 Low-priority data. This data may be forwarded by a Lower Effort PHB in one domain (like the PHB proposed by Informational [RFC3662]), but using the methods specified in this document will be remarked with DSCP CS0 at a Diffserv-Intercon network interconnection. This has the effect that Low-priority data is treated the same as data sent using the default class. (Note: In a network that implements RFC 2474, Low-priority traffic marked as CS1 would otherwise receive better treatment than traffic using the default class.)

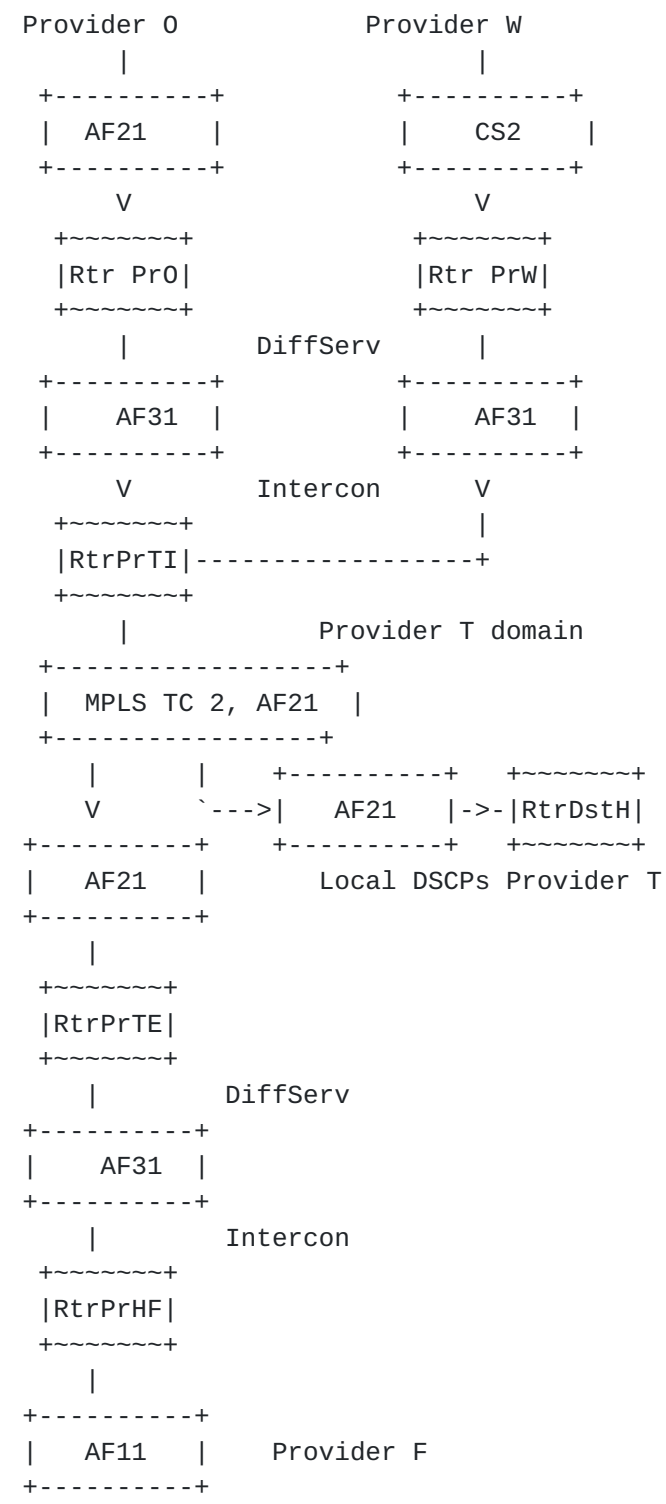
RFC2575 states that Ingress nodes must condition all inbound traffic to ensure that the DS codepoints are acceptable; packets found to have unacceptable codepoints must either be discarded or must have their DS codepoints modified to acceptable values before being forwarded. For example, an ingress node receiving traffic from a domain with which no enhanced service agreement exists may reset the DS codepoint to CS0. As a consequence, an interconnect SLA needs to specify not only the treatment of traffic that arrives with a supported interconnect DSCP, but also the treatment of traffic that arrives with unsupported or unexpected DSCPs; remarking to CS0 is a widely deployed behavior.

4.1. Diffserv-Intercon Example

The overall approach to DSCP marking at network interconnections is illustrated by the following example. Provider O and provider W are

peered with provider T. They have agreed upon a QoS interconnection SLA.

Traffic of provider O terminates within provider T's network, while provider W's traffic transits through the network of provider T to provider F. This example assumes that all providers use their own internal PHB and codepoint (DSCP) that correspond to the AF31 PHB in the Diffserv-Intercon Assured Elastic Treatment Aggregate (AF21 and CS2 are used in the example).



Diffserv-Intercon example

Figure 1

Providers only need to deploy mappings of internal DSCPs to/from Diffserv-Intercon DSCPs in order to exchange traffic using the desired PHBs. In the example, provider O has decided that the properties of his internal class AF21 and are best met by the Diffserv-Intercon Assured Elastic Treatment Aggregate, PHB AF31. At the outgoing peering interface connecting provider O with provider T the former's peering router remarks AF21 traffic to AF31. The domain internal PHB of provider T meeting the requirement of Diffserv-Intercon Assured Elastic Treatment Aggregate are from AF2x PHB group. Hence AF31 traffic received at the interconnection with provider T is remarked to AF21 by the peering router of domain T, and domain T has chosen to use MPLS Traffic Class value 2 for this aggregate. At the penultimate MPLS node, the top MPLS label is removed and exposes the IP header marked by the DSCP which has been set at the network ingress. The peering router connecting domain T with domain F classifies the packet by its domain-T-internal DSCP AF21. As the packet leaves domain T on the interface to domain F, this causes the packet's DSCP to be remarked to AF31. The peering router of domain F classifies the packet for domain-F-internal PHB AF11, as this is the PHB with properties matching Diffserv-Intercon's Assured Elastic Treatment Aggregate.

This example can be extended. The figure shows Provider-W using CS2 for traffic that corresponds to Diffserv-Intercon Assured Elastic Treatment Aggregate PHB AF31; that traffic is mapped to AF31 at the Diffserv-Intercon interconnection to Provider-T. In addition, suppose that Provider-O supports a PHB marked by AF22 and this PHB is supposed to be transported by QoS within Provider-T domain. Then Provider-O will remark it with DSCP AF32 for interconnection to Provider-T.

Finally suppose that Provider-W supports CS3 for internal use only. Then no Diffserv- Intercon DSCP mapping needs to be configured at the peering router. Traffic, sent by Provider-W to Provider-T marked by CS3 due to a misconfiguration may be remarked to CS0 by Provider-T.

4.2. End-to-end QoS: PHB and DS CodePoint Transparency

This section briefly discusses end-to-end QoS approaches related to the Uniform, Pipe and Short Pipe tunnel models ([RFC2983], [RFC3270]), when used edge-to-edge in a network.

- With the Uniform model, neither the DCSP nor the PHB change. This implies that a network management packet received with a CS6 DSCP would be forwarded with an MPLS Traffic Class corresponding to CS6. The uniform model is outside the scope of this document.

- With the Pipe model, the inner tunnel DCSP remains unchanged, but an outer tunnel DSCP and the PHB could changed. For example a packet received with a (network specific) CS1 DSCP would be transported by default PHB and if MPLS is applicable, forwarded with an MPLS Traffic Class corresponding to Default PHB. The CS1 DSCP is not rewritten. Transport of a large variety (much greater than 4) DSCPs may be required across an interconnected network operating MPLS Short pipe transport for IP traffic. In that case, a tunnel based on the Pipe model is among the possible approaches. The Pipe model is outside the scope of this document.

- With the Short Pipe model, the DCSP likely changes and the PHB might change. This document describes a method to simplify QoS for network interconnection when a DSCP rewrite can't be avoided.

4.3. Treatment of Network Control traffic at carrier interconnection interfaces

As specified by RFC4594, section 3.2, Network Control (NC) traffic marked by CS6 is expected at some interconnection interfaces. This document does not change RFC4594, but observes that network control traffic received at network ingress is generally different from network control traffic within a network that is the primary use of CS6 envisioned by RFC 4594. A specific example is that some CS6 traffic exchanged across carrier interconnections is terminated at the network ingress node, e.g. when BGP is used between the two routers on opposite ends of an interconnection link; in this case the operators would enter into a bilateral agreement to use CS6 for that BGP traffic.

The end-to-end QoS discussion in the previous section (4.2) is generally inapplicable to network control traffic - network control traffic is generally intended to control a network, not be transported across it. One exception is that network control traffic makes sense for a purchased transit agreement, and preservation of the CS6 DSCP marking for network control traffic that is transited is reasonable in some cases, although it is generally inappropriate to use CS6 for forwarding that traffic within the network that provides transit. Use of an IP tunnel is suggested in order to conceal the CS6 markings on transiting network control traffic from the network that provides the transit. In this case, Pipe model for Diffserv tunneling is used.

If the MPLS Short Pipe model is deployed for non-tunneled IPv4 traffic, an IP network provider should limit access to the CS6 and CS7 DSCPs so that they are only used for network control traffic for the provider's own network.

Interconnecting carriers should specify treatment of CS6 marked traffic received at a carrier interconnection which is to be forwarded beyond the ingress node. An SLA covering the following cases is recommended when a provider wishes to send CS6 marked traffic across an interconnection link and that traffic's destination is beyond the interconnected ingress node:

- classification of traffic that is network control traffic for both domains. This traffic should be classified and marked for the CS6 DSCP.

- classification of traffic that is network control traffic for the sending domain only. This traffic should be forwarded with a PHB that is appropriate for the NC service class [RFC4594], e.g. AF31 as specified by this document. As an example GSMA IR.34 recommends an Interactive class / AF31 to carry SIP and DIAMETER traffic. While this is service control traffic of high importance to interconnected Mobile Network Operators, it is certainly not Network Control traffic for a fixed network providing transit among such operators, and hence should not receive CS6 treatment in such a transit network.

- any other CS6 marked traffic should be remarked or dropped.

5. Acknowledgements

Bob Briscoe and Gorry Fairhurst reviewed the draft and provided rich feedback. Fred Baker, Brian Carpenter, Al Morton and Sebastien Jobert discussed the draft and helped improving it. Mohamed Boucadair and Thomas Knoll helped adding awareness of related work. James Polk's discussion during IETF 89 helped to improve the text on the relation of this draft to RFC 4594 and RFC 5127.

6. IANA Considerations

This memo includes no request to IANA.

7. Security Considerations

This document does not introduce new features; it describes how to use existing ones. The Diffserv security considerations in RFC 2475 [RFC2475] and RFC 4594 [RFC4594] apply.

8. References

8.1. Normative References

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <http://www.rfc-editor.org/info/rfc2474>.

[RFC3246] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, DOI 10.17487/RFC3246, March 2002, <http://www.rfc-editor.org/info/rfc3246>.

[RFC3270] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, DOI 10.17487/RFC3270, May 2002, <http://www.rfc-editor.org/info/rfc3270>.

[RFC5129] Davie, B., Briscoe, B., and J. Tay, "Explicit Congestion Marking in MPLS", RFC 5129, DOI 10.17487/RFC5129, January 2008, <http://www.rfc-editor.org/info/rfc5129>.

[RFC5865] Baker, F., Polk, J., and M. Dolly, "A Differentiated Services Code Point (DSCP) for Capacity-Admitted Traffic", RFC 5865, DOI 10.17487/RFC5865, May 2010, <http://www.rfc-editor.org/info/rfc5865>.

8.2. Informative References

[I-D.knoll-idr-cos-interconnect] Knoll, T., "BGP Class of Service Interconnection", draft-knoll-idr-cos-interconnect-16 (work in progress), May 2016.

[ID.ietf-idr-sla] IETF, "Inter-domain SLA Exchange", IETF, http://datatracker.ietf.org/doc/draft-ietf-idr-sla-exchange/, 2013.

[IR.34] GSMA Association, "IR.34 Inter-Service Provider IP Backbone Guidelines Version 7.0", GSMA, GSMA IR.34 http://www.gsma.com/newsroom/wp-content/uploads/2012/03/ir.34.pdf, 2012.

[MEF23.1] MEF, "Implementation Agreement MEF 23.1 Carrier Ethernet Class of Service Phase 2", MEF, MEF23.1 http://metroethernetforum.org/PDF_Documents/technical-specifications/MEF_23.1.pdf, 2012.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <http://www.rfc-editor.org/info/rfc2475>.

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, DOI 10.17487/RFC2983, October 2000, <http://www.rfc-editor.org/info/rfc2983>.

[RFC3662] Bless, R., Nichols, K., and K. Wehrle, "A Lower Effort Per-Domain Behavior (PDB) for Differentiated Services", RFC 3662, DOI 10.17487/RFC3662, December 2003, <http://www.rfc-editor.org/info/rfc3662>.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, DOI 10.17487/RFC4594, August 2006, <http://www.rfc-editor.org/info/rfc4594>.

[RFC5127] Chan, K., Babiarz, J., and F. Baker, "Aggregation of Diffserv Service Classes", RFC 5127, DOI 10.17487/RFC5127, February 2008, <http://www.rfc-editor.org/info/rfc5127>.

[RFC5160] Levis, P. and M. Boucadair, "Considerations of Provider-to-Provider Agreements for Internet-Scale Quality of Service (QoS)", RFC 5160, DOI 10.17487/RFC5160, March 2008, <http://www.rfc-editor.org/info/rfc5160>.

[Y.1566] ITU-T, "Quality of service mapping and interconnection between Ethernet, IP and multiprotocol label switching networks", ITU, http://www.itu.int/rec/T-REC-Y.1566-201207-I/en, 2012.

Appendix A. Appendix A The MPLS Short Pipe Model and IP traffic

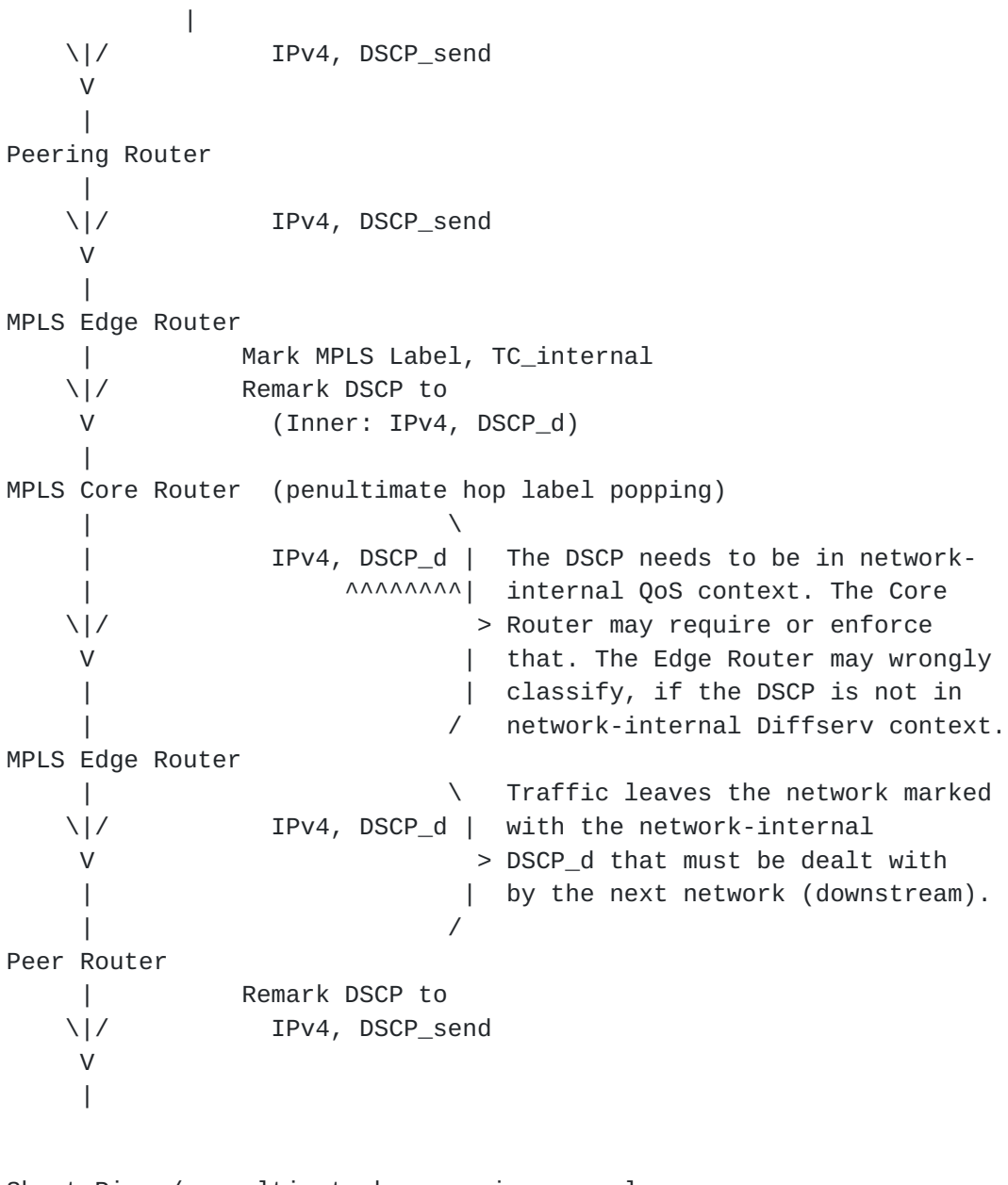
The MPLS Short Pipe Model (or penultimate Hop Label Popping) is widely deployed in carrier networks. If non-tunneled IPv4 traffic is transported using MPLS Short Pipe, IP headers appear inside the last section of the MPLS domain. This impacts the number of PHBs and DSCPs that a network provider can reasonably support. See Figure 2 (below) for an example.

For tunneled IPv4 traffic, only the outer tunnel header is relevant for forwarding. If the tunnel does not terminate within the MPLS network section, only the outer tunnel DSCP is involved, as the inner DSCP does not affect forwarding behavior; in this case all DSCPs could be used in the inner IP header without affecting network behavior based on the outer MPLS header. Here the Pipe model applies.

Non-tunneled IPv6 traffic as well as Layer 2 and Layer 3 VPN traffic all use an additional MPLS label; in this case, the MPLS tunnel follows the Pipe model. Classification and queuing within an MPLS network is always based on an MPLS label, as opposed to the outer IP header.

Carriers often select QoS PHBs and DSCP without regard to interconnection. As a result PHBs and DSCPs typically differ between network carriers. With the exception of best effort traffic, a DSCP change should be expected at an interconnection at least for plain IP traffic, even if the PHB is suitably mapped by the carriers involved.

Although RFC3270 suggests that the Short Pipe Model is only applicable to VPNs, current networks also use it to transport non-tunneled IPv4 traffic. This is shown in figure 2 where Diffserv-Intercon is not used, resulting in exposure of the internal DSCPs of the upstream network to the downstream network across the interconnection.



Short-Pipe / penultimate hop popping example

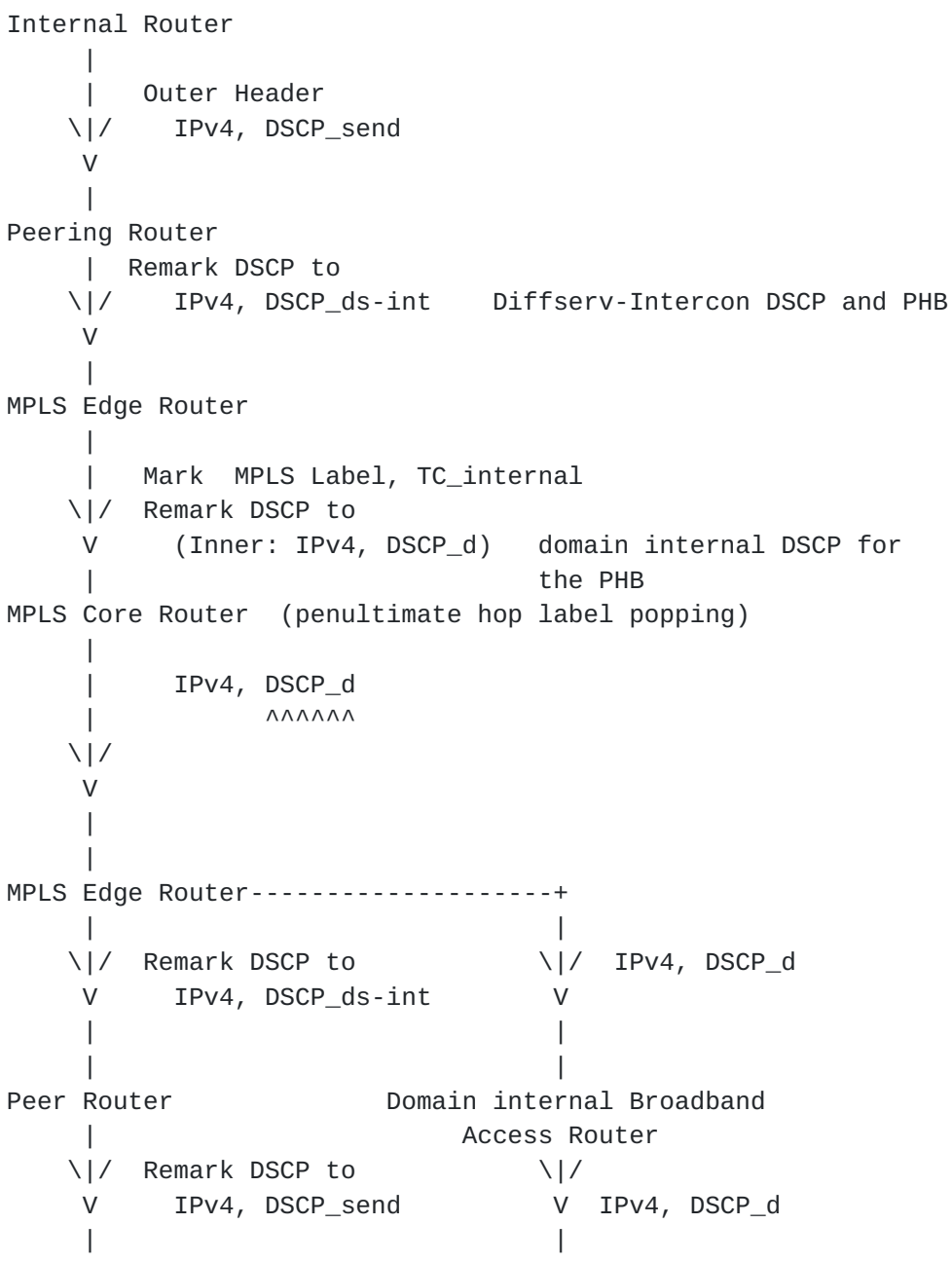
Figure 2

The packets IP DSCP must be in a well understood Diffserv context for schedulers and classifiers on the interfaces of the ultimate MPLS link (last link traversed before leaving the network). The necessary Diffserv context is network-internal and a network operating in this mode enforces DSCP usage in order to obtain robust QoS behavior.

Without Diffserv-Intercon treatment, the traffic is likely to leave each network marked with network-internal DSCP. DSCP_send in the figure above has to be remarked into the first network's Diffserv

scheme at the ingress MPLS Edge Router, to DSCP_d in the example. For that reason, the traffic leaves this domain marked by the network-internal DSCP_d. This structure requires that every carrier deploys per-peer PHB and DSCP mapping schemes.

If Diffserv-Intercon is applied DSCPs for traffic transiting the domain can be mapped from and remapped to an original DSCP. This is shown in figure 3. Internal traffic may continue to use internal DSCPs (e.g, DSCP_d) and they may also be used between a carrier and its direct customers.



Short-Pipe example with Diffserv-Intercon

Figure 3

Appendix B. Change log (to be removed by the RFC editor)

00 to 01 Added an Applicability Statement. Put the main part of the RFC5127 related discussion into a separate chapter.

01 to 02 More emphasis on the Short-Pipe tunel model as compared to Pipe and Uniform tunnel models. Further editorial improvements.

02 to 03 Suggestions how to remark all RFC4594 classes to Diffserv-Intercon classes at interconnection.

03 to 04 Minor clarifications and editorial review, preparation for WGLC.

Authors' Addresses

Ruediger Geib (editor)
Deutsche Telekom
Heinrich Hertz Str. 3-7
Darmstadt 64295
Germany

Phone: +49 6151 5812747
Email: Ruediger.Geib@telekom.de

David L. Black
EMC Corporation
176 South Street
Hopkinton, MA
USA

Phone: +1 (508) 293-7953
Email: david.black@emc.com