Internet Engineering Task Force (IETF)
Request for Comments: 8100
Category: Informational
ISSN: 2070-1721

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

March 2017

Diffserv-Interconnection Classes and Practice

Abstract

This document defines a limited common set of Diffserv Per-Hop Behaviors (PHBs) and Diffserv Codepoints (DSCPs) to be applied at (inter)connections of two separately administered and operated networks, and it explains how this approach can simplify network configuration and operation. Many network providers operate Multiprotocol Label Switching (MPLS) using Treatment Aggregates for traffic marked with different Diffserv Per-Hop Behaviors 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 Model. While motivated by the requirements of MPLS network operators that use Short Pipe Model tunnels, this document is applicable to other networks, both MPLS and non-MPLS.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc8100.

Table of Contents

1. Introduction

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

As described by Section 2.3.4.2 of [RFC2475], the re-marking of packets at domain boundaries is a Diffserv feature. 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 re-mark unknown or unexpected DSCPs to zero upon receipt for default (best-effort) forwarding in accordance with the guidance in [RFC2475] to ensure that appropriate DSCPs are used within a Diffserv domain. This document is based on the latter approach and defines additional DSCPs that are known and expected at network interconnection interfaces in order to reduce the amount of traffic whose DSCPs are re-marked to zero.

This document is motivated by requirements for IP network interconnection with Diffserv support among providers that operate Multiprotocol Label Switching (MPLS) in their backbones, but it is also 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 Model for MPLS tunnel 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 Diffserv 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 Diffserv treatment is more likely to result when an interconnection codepoint scheme is used because traffic is re-marked to the same DSCPs at all network interconnections.

The interconnection approach described in this document (referred to as "Diffserv-Intercon") uses a set of PHBs (mapped to four corresponding 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 benefiting from a defined Diffserv interconnection interface.

The basic idea is that traffic sent with a Diffserv-Intercon PHB and DSCP is restored to that PHB and DSCP at each network interconnection, even though a different PHB and DSCP may be used within each network involved. The key requirement is that the network ingress interconnect DSCP be restored at the 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 re-marked to an interconnect DSCP or dealt with via an additional agreement(s) among the operators of the interconnected networks; use of the MPLS Short Pipe Model favors re-marking unexpected DSCPs to zero in the absence of an additional agreement(s), as explained further 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. [RFC5160] suggests Meta-QoS-Classes to help enable deployment of standardized end-to-end QoS classes. The Diffserv-Intercon class and codepoint scheme is intended to complement that work (e.g., by enabling a defined set of interconnection DSCPs and PHBs).

Border Gateway Protocol (BGP) support for signaling Class of Service at interconnection interfaces [BGP-INTERCONNECTION] [SLA-EXCHANGE] is complementary to Diffserv-Intercon. These two BGP documents focus on exchanging Service Level Agreement (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 PHB id codes.

1.2. Applicability Statement

This document is applicable to the use of Differentiated Services for interconnection traffic between networks and is particularly suited to interconnection of MPLS-based networks that use MPLS Short Pipe Model tunnels. This document is also applicable to other network technologies, but it is not intended for use within an individual network, where the approach specified in [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 is also applicable to Pipe Model tunneling [RFC2983] [RFC3270], but it is not applicable to Uniform Model tunneling [RFC2983] [RFC3270].

The Diffserv-Intercon approach defines a set of four PHBs for support at interconnections (or network boundaries in general). Corresponding DSCPs for use at an interconnection interface are also defined. Diffserv-Intercon allows for a simple mapping of PHBs and DSCPs to MPLS Treatment Aggregates. It is extensible by IETF standardization, and this allows additional PHBs and DSCPs to be specified for the Diffserv-Intercon scheme. Coding space for private interconnection agreements or provider internal services is available, as only a single digit number of standard DSCPs are applied by the Diffserv-Intercon approach.

1.3. Document Organization

This document is organized as follows: Section 2 reviews the MPLS Short Pipe Model for Diffserv Tunnels [RFC3270], because effective support for that model is a crucial goal of Diffserv-Intercon. Section 3 provides background on the approach described in RFC 5127 to Traffic Class (TC) aggregation within a Diffserv network domain and contrasts it with the Diffserv-Intercon approach. Section 4 introduces Diffserv-Intercon 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 IP traffic, MPLS VPN Diffserv considerations, and the handling of high-priority network management traffic. Appendix A describes how the MPLS Short Pipe Model (Penultimate Hop Popping (PHP)) impacts DSCP marking for IP interconnections.

2. MPLS and Short Pipe Model Tunnels

This section provides a summary of the implications of MPLS Short Pipe Model tunnels and, in particular, their use of PHP (see RFC 3270) on the Diffserv tunnel framework described in RFC 2983. The Pipe and Uniform Models for Differentiated Services and Tunnels are defined in [RFC2983]. RFC 3270 adds the Short Pipe Model to reflect the impact of MPLS PHP, 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 unencapsulated IP traffic. This has important implications for Diffserv functionality in MPLS networks.

Per RFC 2474, the recommendation to forward traffic with unrecognized DSCPs with default (best-effort) service without rewriting the DSCP has not been widely deployed in 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 re-mark 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 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 Pipe Model tunnels.

Short Pipe Model tunnels 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 information being available at the 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 the 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 the approach to aggregation of Diffserv TCs for use within a network as described in RFC 5127, but there are important differences caused by characteristics of network interconnects that differ from links within a network.

3.1. Background of RFC 5127

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 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 TCs.

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

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

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

- Migrations from one DSCP scheme to a different one is another possible application of otherwise unused DSCPs.

3.2. Differences from RFC 5127

Like RFC 5127, this document also uses four Treatment Aggregates, but it differs from RFC 5127 in some important ways:

- It follows RFC 2475 in allowing the DSCPs used within a network to differ from those used to exchange traffic with other networks (at network edges), but it provides support to restore ingress DSCP values if one of the recommended interconnect DSCPs in this document is used. This results in DSCP re-marking at both network ingress and network egress, and this document assumes that such re-marking 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 Aggregate 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 (NC) traffic as a special case. Within a provider's network, the CS6 DSCP is used for local network control traffic (routing protocols and Operations, Administration, and Maintenance (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 Treatment Aggregate for use at network interconnections instead of the Network Control Treatment 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 Treatment Aggregate for Voice over IP (VoIP) traffic, where network-provided service differentiation 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 four interconnection Treatment Aggregates with well-defined DSCPs to be aggregated by them. A sending party re-marks DSCPs from internal usage to the interconnection codepoints. The receiving party re-marks DSCPs to their internal usage. 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 a small number of Treatment 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], Global System for Mobile Communications Association (GSMA) IR.34 [IR.34], and MEF23.1 [MEF23.1].

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

Telephony Service Treatment Aggregate: PHB Expedited Forwarding (EF), DSCP 101 110 and PHB VOICE-ADMIT, DSCP 101 100 (see [RFC3246], [RFC4594], and [RFC5865]). This Treatment Aggregate corresponds to the Real-Time Treatment Aggregate definition regarding the queuing (both delay and jitter should be minimized) per RFC 5127, but this aggregate is restricted to transport Telephony service class traffic in the sense of [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 to provide Diffserv-Intercon network interconnection of a subset of the Real-Time Treatment Aggregate defined in RFC 5127, specifically the portions that consume significant bandwidth. This traffic is expected to consist of the following classes defined in RFC 4594: Broadcast Video, Real-Time Interactive, and Multimedia Conferencing. This Treatment Aggregate should be configured with a rate-based queue (consistent with the recommendation for the transported TCs in RFC 4594). 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 Conferencing traffic.

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 the extension of PHBs to be aggregated by this Treatment Aggregate. For Diffserv-Intercon network interconnection, the following service

classes (per RFC 4594) should be mapped to the Assured Elastic Treatment Aggregate: the Signaling service class (being marked for lowest loss probability), the 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. An example in RFC 5127 refers to this Treatment Aggregate as "Elastic Treatment Aggregate". An important difference from RFC 5127 is that any traffic with unrecognized or unsupported DSCPs may be re-marked to this DSCP. For Diffserv-Intercon network interconnection, the Standard service class and Low-Priority Data service class defined in RFC 4594 should be mapped to this Treatment Aggregate. This document does not specify an interconnection class for Low-Priority Data (also defined RFC 4594). This traffic may be forwarded with a Lower Effort PHB in one domain (e.g., the PHB proposed by Informational [RFC3662]), but the methods specified in this document re-mark this traffic 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 Standard service class. (Note: In a network that implements RFC 2474, Low-Priority traffic marked as CS1 would otherwise receive better treatment than Standard traffic using the default PHB.)

RFC 2475 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 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; re-marking to CS0 is a widely deployed behavior.

During the process of setting up a Diffserv interconnection, both networks should define the set of acceptable and unacceptable DSCPs and specify the treatment of traffic marked with each DSCP.

While Diffserv-Intercon allows modification of unacceptable DSCPs, if traffic using one or more of the PHBs in a PHB group (e.g., AF3x, consisting of AF31, AF32, and AF33) is accepted as part of a supported Diffserv-Intercon Treatment Aggregate, then traffic using other PHBs from the same PHB group should not be modified to use PHBs outside of that PHB group and, in particular, should not be re-marked

to CS0 unless the entire PHB group is re-marked to CS0. This avoids unexpected forwarding behavior (and potential reordering; see also [RFC7657]) when using Assured Forwarding (AF) PHBs [RFC2597].

4.1. Diffserv-Intercon Example

The overall approach to DSCP marking at network interconnections is illustrated by the following example. Provider O, provider W, and provider F are peered with provider T. They have agreed upon a Diffserv 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, CS2, and AF11 are used in the example).

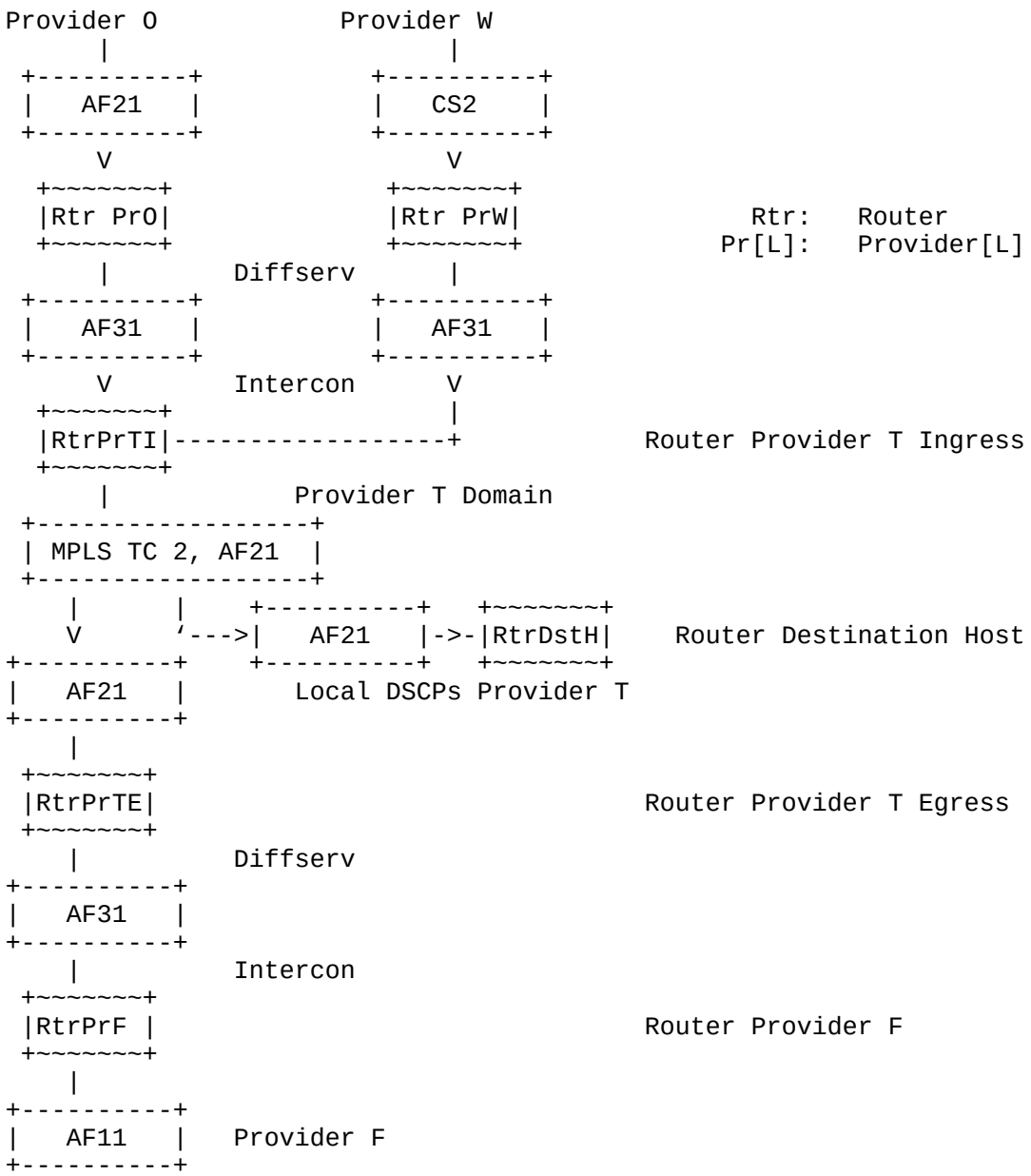

Figure 1: Diffserv-Intercon Example

Providers only need to deploy mappings of internal DSCPs to/from Diffserv-Intercon DSCPs, so that they can exchange traffic using the desired PHBs. In the example, provider O has decided that the properties of his internal class AF21 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 re-marks AF21 traffic to AF31. The domain internal PHB of provider T that meets the requirement of the Diffserv-Intercon Assured Elastic Treatment Aggregate is from the AF2x PHB group. Hence, AF31 traffic received at the interconnection with provider T is re-marked to AF21 by the peering router of domain T, and domain T has chosen to use MPLS TC 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 that 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 re-marked 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 the Diffserv-Intercon 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 obtain Diffserv transport within provider T's domain. Then provider O will re-mark 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 re-marked to CS0 by provider T.

4.2. End-to-End PHB and DSCP Transparency

This section briefly discusses end-to-end Diffserv approaches related to the Uniform, Pipe, and Short Pipe Model tunnels [RFC2983] [RFC3270] when used edge-to-edge in a network.

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

- With the Pipe Model, the inner tunnel DSCP remains unchanged, but an outer tunnel DSCP and the PHB could change. For example, a packet received with a (network-specific) CS1 DSCP would be transported by a Default PHB and, if MPLS is applicable, forwarded with an MPLS TC corresponding to the Default PHB. The CS1 DSCP is not rewritten. Transport of a large variety (much greater than four) DSCPs may be required across an interconnected network operating MPLS Short Pipe Model 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 DSCP likely changes, and the PHB might change. This document describes a method to simplify Diffserv network interconnection when a DSCP rewrite can’t be avoided.

4.3. Treatment of Network Control Traffic at Carrier Interconnection Interfaces

As specified in Section 3.2 of RFC 4594, NC traffic marked by CS6 is expected at some interconnection interfaces. This document does not change RFC 4594 but observes that network control traffic received at a 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 discussion in Section 4.2 is generally inapplicable to network control traffic -- network control traffic is generally intended to control a network, not be transported between networks. 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, the Pipe Model for Diffserv tunneling is used.

If the MPLS Short Pipe Model is deployed for unencapsulated 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 that 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 transiting NC service class traffic [RFC4594] in the receiving domain, 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 re-marked or dropped.

5. IANA Considerations

This document does not require any IANA actions.

6. Security Considerations

The DSCP field in the IP header can expose additional traffic classification information at network interconnections by comparison to the use of a zero DSCP for all interconnect traffic. If traffic classification information is sensitive, the DSCP field could be re-marked to zero to hide the classification as a countermeasure, at the cost of loss of Diffserv information and differentiated traffic handling on the interconnect and subsequent networks. When AF PHBs are used, any such re-marking should respect AF PHB group boundaries as further discussed at the end of Section 4.

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

7. References

7.1. 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>.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, DOI 10.17487/RFC2597, June 1999, <http://www.rfc-editor.org/info/rfc2597>.

[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>.

7.2. Informative References

[BGP-INTERCONNECTION] Knoll, T., "BGP Class of Service Interconnection", Work in Progress, draft-knoll-idr-cos-interconnect-17, November 2016.

[IR.34] GSMA, "Guidelines for IPX Provider networks (Previously Inter-Service Provider IP Backbone Guidelines)", Official Document IR.34, Version 11.0, November 2014, <http://www.gsma.com/newsroom/wp-content/uploads/IR.34-v11.0.pdf>.

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

[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>.

[RFC7657] Black, D., Ed. and P. Jones, "Differentiated Services (Diffserv) and Real-Time Communication", RFC 7657, DOI 10.17487/RFC7657, November 2015, <http://www.rfc-editor.org/info/rfc7657>.

[SLA-EXCHANGE] Shah, S., Patel, K., Bajaj, S., Tomotaki, L., and M. Boucadair, "Inter-domain SLA Exchange Attribute", Work in Progress, draft-ietf-idr-sla-exchange-10, January 2017.

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

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 unencapsulated IP 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 for an example.

For encapsulated IP 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.

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 PHBs and DSCPs 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 unencapsulated IP traffic, even if the PHB is suitably mapped by the carriers involved.

Although RFC 3270 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.

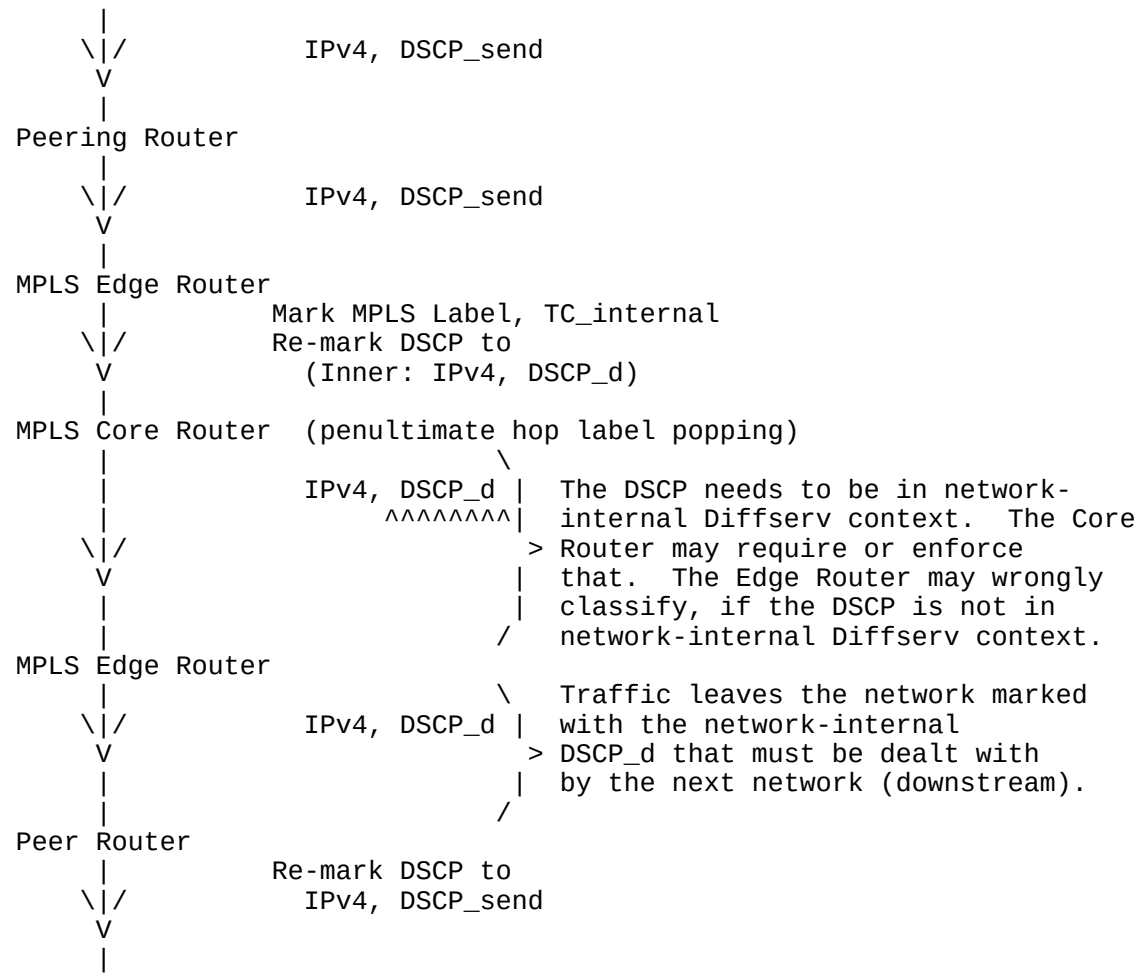

Figure 2: Short Pipe Model / Penultimate Hop Popping Example

The packet's 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 differentiated forwarding 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 re-marked 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.

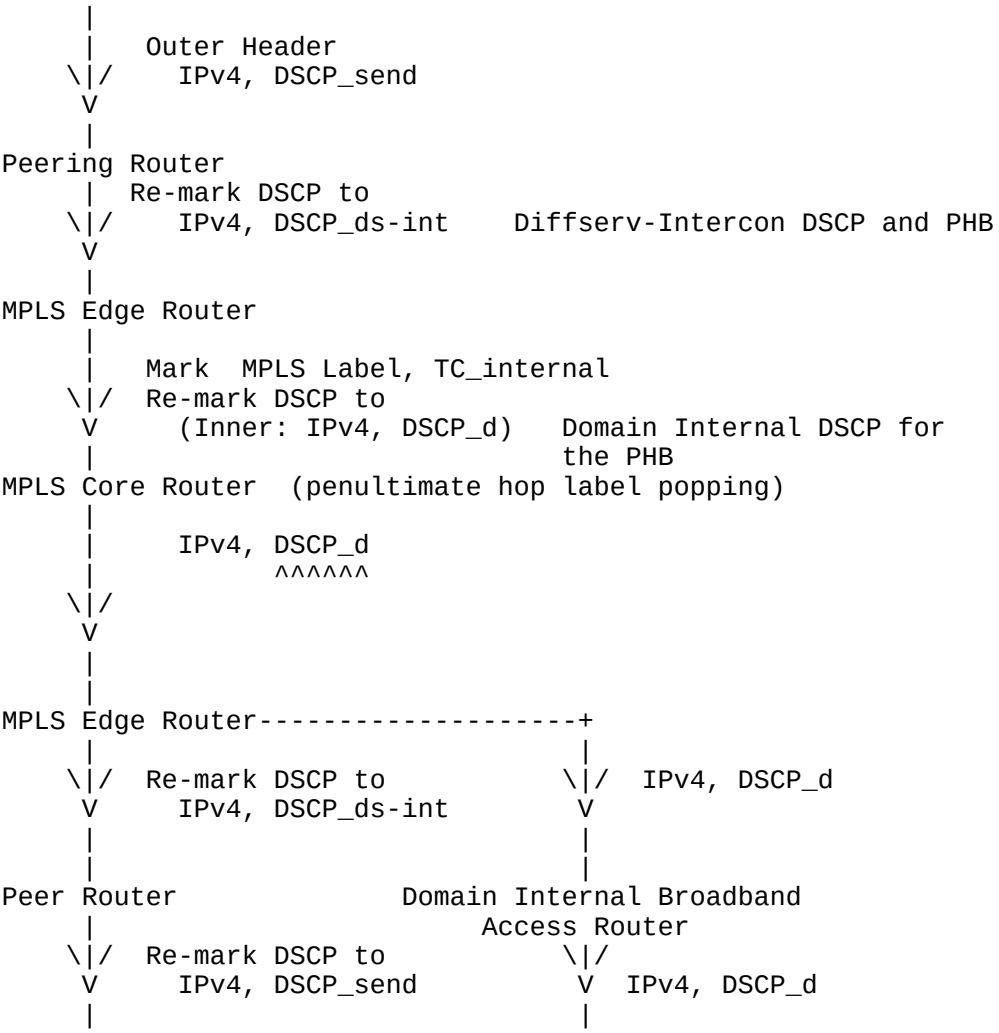

Figure 3: Short Pipe Model Example with Diffserv-Intercon

Acknowledgements

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

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
Dell EMC
176 South Street
Hopkinton, MA
United States of America

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