

Tsvwg Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: September 14, 2016

J. You  
Huawei  
M. Welzl  
University of Oslo  
B. Trammell  
M. Kuehlewind  
ETH Zurich  
K. Smith  
Vodafone Group  
March 13, 2016

**Latency Loss Tradeoff PHB Group**  
**draft-you-tsvwg-latency-loss-tradeoff-00**

Abstract

This document defines a PHB (Per-Hop Behavior) group called Latency Loss Tradeoff (LLT). The LLT group is intended to provide delivery of IP packets in two classes of services: a low-loss service (Lo service) and a low-latency service (La service). The LLT group enables an application to request treatment for either low-loss or low-latency at a congested network link.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

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 September 14, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- [1.](#) Introduction . . . . . [3](#)
- [2.](#) Terminology . . . . . [3](#)
  - [2.1.](#) Abbreviations and acronyms . . . . . [3](#)
  - [2.2.](#) Definitions . . . . . [4](#)
- [3.](#) Problem Statement . . . . . [5](#)
  - [3.1.](#) Existing DSCP PHBs . . . . . [5](#)
    - [3.1.1.](#) Default PHB . . . . . [5](#)
    - [3.1.2.](#) Class Selector PHB . . . . . [5](#)
    - [3.1.3.](#) Assured Forwarding PHB Group . . . . . [5](#)
    - [3.1.4.](#) Expedited Forwarding PHB . . . . . [6](#)
    - [3.1.5.](#) Voice Admit PHB . . . . . [6](#)
    - [3.1.6.](#) Delay Bound PHB . . . . . [6](#)
  - [3.2.](#) Incentives . . . . . [6](#)
- [4.](#) Definition of LLT PHB . . . . . [7](#)
  - [4.1.](#) Goal and Scope of LLT . . . . . [7](#)
  - [4.2.](#) Description of LLT behavior . . . . . [8](#)
    - [4.2.1.](#) Implementation Considerations . . . . . [8](#)
  - [4.3.](#) Microflow misordering . . . . . [9](#)
  - [4.4.](#) Recommended Codepoints . . . . . [9](#)
  - [4.5.](#) Mutability . . . . . [9](#)
  - [4.6.](#) Tunneling . . . . . [9](#)
  - [4.7.](#) Interaction with other PHBs . . . . . [9](#)
- [5.](#) Security Considerations . . . . . [10](#)
- [6.](#) IANA Considerations . . . . . [10](#)
- [7.](#) References . . . . . [10](#)
  - [7.1.](#) Normative References . . . . . [10](#)
  - [7.2.](#) Informative References . . . . . [11](#)
- Authors' Addresses . . . . . [11](#)



## **1. Introduction**

Different applications have different communication requirements [[QoS](#)]. In interactive applications of real-time sound transmission, as well as in virtual reality, the overall one-way delay needs to be short in order to give the user an impression of a real-time response. Yet, these applications may be able to tolerate high loss rates. In conventional text and data networking, delay thresholds are the least stringent. The response time in these types of applications can increase from 2 to 5 seconds before becoming unacceptable. However, given that increased loss reduces the throughput of TCP, these applications desire minimal loss.

The network resources consist primarily of buffers and link bandwidth. Operators often favor high utilization of bottleneck links at the price of high queuing delay. This is beneficial for non-real time applications. However, this may be considered unacceptable for some real-time applications. The proposed LLT group enables an application to choose between low-latency and low-loss at a congested network link [[ABE](#)] [[RD](#)]. Typically, an interactive application with real-time deadlines, such as audio, will mark most of its packets as a low-latency service. In contrast, an application that transfers bulk data will mark most of its packets as a low-loss service. The LLT group can be thought of as allowing an application to trade loss for delay by marking packets low-latency service (La) or to trade delay for loss by marking packets low-loss service (Lo).

## **2. Terminology**

This section contains definitions for terms used frequently throughout this document.

### **2.1. Abbreviations and acronyms**

DS: Differentiated Service

PHB: Per-Hop Behavior

LLT: Latency Loss Tradeoff

TCA: Traffic Conditioning Agreement

TCP: Transmission Control Protocol



## **2.2. Definitions**

DS-capable: capable of implementing differentiated services as described in this architecture; usually used in reference to a domain consisting of DS-compliant nodes.

DS codepoint: a specific value of the DSCP portion of the DS field, used to select a PHB.

DS-compliant: enabled to support differentiated services functions and behaviors as defined in [[RFC2474](#)], this document, and other differentiated services documents; usually used in reference to a node or device.

DS field: the IPv4 header TOS octet or the IPv6 Traffic Class octet when interpreted in conformance with the definition given in [[RFC2474](#)]. The bits of the DSCP field encode the DS codepoint, while the remaining bits are currently unused.

Low-latency service (La service): puts an emphasis on low queuing delay at a congested network link. It allows an application to trade loss for delay.

Low-loss service (Lo service): puts an emphasis on low packet loss rate at a congested network link. It allows an application to trade delay for loss.

Per-Hop-Behavior (PHB): the externally observable forwarding behavior applied at a DS-compliant node to a DS behavior aggregate.

PHB group: a set of one or more PHBs that can only be meaningfully specified and implemented simultaneously, due to a common constraint applying to all PHBs in the set such as a queue servicing or queue management policy. A PHB group provides a service building block that allows a set of related forwarding behaviors to be specified together (e.g., four dropping priorities). A single PHB is a special case of a PHB group.

Traffic Conditioning Agreement (TCA): an agreement specifying classifier rules and any corresponding traffic profiles and metering, marking, discarding and/or shaping rules which are to apply to the traffic streams selected by the classifier. A TCA encompasses all of the traffic conditioning rules explicitly specified within a SLA along with all of the rules implicit from the relevant service requirements and/or from a DS domain's service provisioning policy.



**3. Problem Statement**

**3.1. Existing DSCP PHBs**

**3.1.1. Default PHB**

A default Per-Hop Behavior (PHB) [[RFC2474](#)] MUST be available in a DiffServ (DS)-compliant node. This is the common, best-effort forwarding behavior available in existing routers as standardized in [[RFC1812](#)]. Codepoint '000000' from Pool 1 is used as the default PHB value. In this document, packets received with the Default PHB is treated as Lo service on the LLT-compliant router.

**3.1.2. Class Selector PHB**

The Class Selector (CS) PHB [[RFC2474](#)] is introduced for backwards compatibility with use of the IPv4 Precedence field. Any of the eight codepoints in the range 'xxx000' (where 'x' may equal '0' or '1') from Pool 1 is assigned as Class Selector codepoint. The CS PHB does not match the services that LLT PHB is trying to deliver.

**3.1.3. Assured Forwarding PHB Group**

The Assured Forwarding (AF) PHB group [[RFC2597](#)] allows the operator to provide assured forwarding of IP packets as long as the aggregate traffic does not exceed the subscribed rate. Traffic that exceeds the subscribed rate is not delivered with as high a probability as the traffic that is within the rate.

The AF PHB group provides delivery of IP packets in four independently forwarded AF classes. Within each AF class, an IP packet can be assigned one of three different levels of drop precedence. The combination of classes and drop precedence yields twelve separate DSCP encodings from AF11 through AF43 as follows:

	Class 1	Class 2	Class 3	Class 4
Low Drop Prec	001010	010010	011010	100010
Medium Drop Prec	001100	010100	011100	100100
High Drop Prec	001110	010110	011110	100110

The AF PHB does not match the services that LLT PHB is trying to deliver.





#### **3.1.4. Expedited Forwarding PHB**

Expedited Forwarding (EF) PHB [[RFC3246](#)] is intended to provide a building block for low delay, low jitter and low loss services by ensuring that the EF aggregate is served at a certain configured rate. EF traffic requires a strict admission control mechanism. Codepoint '101110' is recommended for the EF PHB. The EF PHB does not match the services that LLT PHB is trying to deliver.

#### **3.1.5. Voice Admit PHB**

The Voice Admit (VA) PHB [[RFC5865](#)] has identical characteristics to the Expedited Forwarding PHB. However Voice Admit traffic is also admitted by the network using a Call Admission Control (CAC) procedure. The recommended DSCP for Voice Admit is '101100', parallel with the existing EF codepoint '101110'. The VA PHB does not match the services that LLT PHB is trying to deliver.

#### **3.1.6. Delay Bound PHB**

The Delay Bound (DB) PHB [[RFC3248](#)] requires a bound on the delay of packets due to other traffic in the network. Two parameters - capped arrival rate (R) and a 'score' (S), are defined and related to the target delay variation bound. An experimental codepoint '101111' is suggested for DB behavior. In this document, there's no specific bound on the delay, the LLT PHB only indicates the tradeoff.

### **3.2. Incentives**

The primary goal of differentiated services is to allow different levels of service to be provided for traffic streams on a common network infrastructure. Hence, an adversary may be able to obtain better service by modifying the DS field to codepoints indicating behaviors used for enhanced services or by injecting packets with the DS field set to such codepoints. Such theft-of-service ([RFC2474](#)), [RFC2475](#)) becomes a denial-of-service attack when the modified or injected traffic depletes the resources available to forward it and other traffic streams.

DS ingress nodes must condition all traffic entering a DS domain to ensure that it has acceptable DS codepoints. This means that the codepoints must conform to the applicable TCA(s) (Traffic Conditioning Agreement) [[RFC2475](#)] and the domain's service provisioning policy. Packets received with an 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 the



Default PHB codepoint. However, the Default PHB (i.e. best-effort forwarding) cannot meet the diverse needs of different Internet applications.

The objective of the LLT PHB group is to retain the best-effort service while providing low delay to real-time applications at the expense of increased loss or providing low loss to non real-time applications at the expense of increased delay. This requires Internet applications to mark their traffic with appropriate codepoint values. Since the low-loss service is neither better nor worse than the low-latency service but is merely different, there is no incentive for Internet applications to abuse such codepoints, and no need for admission control.

#### **4. Definition of LLT PHB**

The LLT group provides forwarding of IP packets in two classes of service: a low-loss service (Lo) and a low-latency service (La). The LLT group enables an application to choose between low latency and low loss at a congested network link. The packets marked as low-latency service receive little queuing delay. The packets marked as low-loss service receive at least as much throughput as they would in a legacy best effort network. La-marked packets are more likely to be dropped during periods of congestion than the Lo-marked packets. Note that among the two services, neither of the two has priority over the other.

##### **4.1. Goal and Scope of LLT**

The LLT group may be used by a network operator in two distinct ways: either as a separate service, or as a replacement of the flat (existing) best-effort IP service.

A DS (Differentiated Services) node SHOULD implement the LLT group. It MAY allocate a configurable, minimum amount of forwarding resources (buffer space and bandwidth) to LLT group.

The LLT group MAY also be configurable to receive more forwarding resources than the minimum when excess resources are available from other PHB groups. This is beyond the scope of this document.

The LLT PHB definition does NOT mandate or recommend any particular method for achieving LLT behavior.



## **4.2. Description of LLT behavior**

To support the LLT group on an output link, the router can maintain two FIFO (First-In First-Out) queues referred to as a Lo (Loss-sensitive) queue and La (Latency-sensitive) queue for packets destined to the link. Depending on whether an incoming packet is marked for the low-loss or low-latency service, the router appends the packet to the Lo or La queue respectively. The packets within each queue are served in the FIFO order. The scheduling is work-conserving.

A router can support the desired delay differentiation between the Lo and La services through buffer sizing for the Lo and La queues, and by ensuring that the La queue does not grow larger than the Lo queue. As common in current Internet routers, the size of the Lo buffer is chosen large enough so that the oscillating transmission of TCP (Transmission Control Protocol) and other legacy end-to-end congestion control protocols utilizes the available link rate fully. The La buffer is configured to a much smaller dynamic size to ensure that queuing delay for each forwarded packet of the La class is low. The assurance of low maximum queuing delay is attractive for delay-sensitive applications and easily verifiable by outside parties.

### **4.2.1. Implementation Considerations**

This document does not specify any particular implementation method for achieving LLT behavior. Some LLT-like implementations may refer to [[I-D.hurley-alternative-best-effort](#)], [[RD](#)] and [[I-D.briscoe-aqm-dualq-coupled](#)].

[[I-D.hurley-alternative-best-effort](#)] marks every best effort packet as either green or blue. Green packets receive a low, bounded delay at every hop, the value of the per-hop delay bound configured by the operator. However, when transmitting more aggressively, the green users can enjoy both a higher rate and lower queuing delay than those of the blue users, which weakens the incentives for incremental deployment. [[RD](#)] proposes Rate-Delay (RD) service enabling a user to choose either a higher transmission rate or low queuing delay. The R (Rate) service is like Lo service while D (Delay) service is like La service.

Note that both classes defined in this document do not provide any absolute guarantees on the loss rate or delay a packet will experience. Using these classes only provides a relative treatment compared to the other class. Depending on the amount of traffic arriving per class, it is possible for traffic in the La class to experience more delay than traffic in the Lo class. However, this may be circumvented by using scheduling mechanisms, for example, by



adjusting the scheduling function that assigns traffic to the Lo and La queues, or by adjusting the scheduling weight based on the average load in each class. Moreover, the delay experienced by La traffic is always bounded by the length of the La queue. The particular implementation is beyond the scope of this document.

When a DS-compliant node claims to implement the LLT PHB, the implementation MUST conform to the specification given in this document.

#### **4.3. Microflow misordering**

The packets within each queue are served in the FIFO order. Packets belonging to a single microflow within the LLT aggregate SHOULD NOT experience re-ordering in normal operation of the device when passing through.

#### **4.4. Recommended Codepoints**

Recommended codepoints for the LLT PHB group are given below.

Low-loss service: 000001  
Low-latency service:000101

#### **4.5. Mutability**

Packets marked for LLT PHB MAY be remarked at a DS domain boundary only to other codepoints that satisfy the LLT PHB. Packets marked for LLT PHBs SHOULD NOT be demoted or promoted to another PHB by a DS domain.

#### **4.6. Tunneling**

When LLT packets are tunneled, the tunneling packets must be marked as LLT.

#### **4.7. Interaction with other PHBs**

Other PHBs and PHB groups may be deployed in the same DS node or domain with the LLT PHB.

Packets received with the Default PHB SHOULD be treated as Lo service as default by the LLT PHB aware device. [RD] has proved that La service does not hurt Lo service.

Packets received with the LLT PHB SHOULD be treated as Default PHB as default by the LLT PHB unaware device.





## **5. Security Considerations**

Internet applications cannot benefit from wrongly indicating low-loss or low-latency as they have to pay the expense of high delay or high loss as a tradeoff. Hence there is no incentive for Internet applications to set the wrong codepoints.

## **6. IANA Considerations**

This document suggests two experimental codepoints, 000001/000101, in Pool 3 of the code space defined by [[RFC2474](#)].

## **7. References**

### **7.1. Normative References**

- [RFC1812] Baker, F., Ed., "Requirements for IP Version 4 Routers", [RFC 1812](#), DOI 10.17487/RFC1812, June 1995, <<http://www.rfc-editor.org/info/rfc1812>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [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>>.
- [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>>.
- [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>>.



- [RFC3248] Armitage, G., Carpenter, B., Casati, A., Crowcroft, J., Halpern, J., Kumar, B., and J. Schnizlein, "A Delay Bound alternative revision of [RFC 2598](#)", [RFC 3248](#), DOI 10.17487/RFC3248, March 2002, <<http://www.rfc-editor.org/info/rfc3248>>.
- [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**

- [ABE] Hurley, P., Le Boudec, J., Thiran, P., and M. Kara, "ABE: Providing a Low-Delay Service within Best Effort", IEEE Network Magazine 15(3): 60-69, May 2001.
- [I-D.briscoe-aqm-dualq-coupled]  
Schepper, K., Briscoe, B., Bondarenko, O., and I. Tsang, "DualQ Coupled AQM for Low Latency, Low Loss and Scalable Throughput", [draft-briscoe-aqm-dualq-coupled-00](#) (work in progress), August 2015.
- [I-D.hurley-alternative-best-effort]  
Hurley, P., Iannaccone, G., Kara, M., Le Boudec, J., Thiran, P., and C. Diot, "The ABE Service", November 2000.
- [QoS] Chen, C., Farley, T., and N. Ye, "QoS Requirements of Network Applications on the Internet", Information Knowledge Systems Management 2004, 4(1): 55-76, 2004.
- [RD] Podlesny, M. and S. Gorinsky, "RD network services: differentiation through performance incentives", ACM SIGCOMM Computer Communication Review, 38(4): 255-266, 2008.

### Authors' Addresses

Jianjie You  
Huawei  
101 Software Avenue, Yuhua District  
Nanjing 210012  
China

Email: youjianjie@huawei.com



Michael Welzl  
University of Oslo  
PO Box 1080 Blindern  
Oslo N-0316  
Norway

Email: [michawe@ifi.uio.no](mailto:michawe@ifi.uio.no)

Brian Trammell  
ETH Zurich  
Zurich  
Switzerland

Email: [ietf@trammell.ch](mailto:ietf@trammell.ch)

Mirja Kuehlewind  
ETH Zurich  
Zurich  
Switzerland

Email: [mirja.kuehlewind@tik.ee.ethz.ch](mailto:mirja.kuehlewind@tik.ee.ethz.ch)

Kevin Smith  
Vodafone Group  
One Kingdom Street,  
London  
UK

Email: [kevin.smith@vodafone.com](mailto:kevin.smith@vodafone.com)

