

INTERNET-DRAFT
Expires: March 2000

R. Bless
K. Wehrle
Universitaet Karlsruhe (TH)
September 1999

A Lower Than Best-Effort Per-Hop Behavior

<[draft-bless-diffserv-lbe-phb-00.txt](#)>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

Abstract

This document describes a Lower than Best-Effort (LBE) per-hop behavior (PHB) for use within and between Differentiated Services domains [3]. The primary objective of this LBE PHB is to separate LBE traffic from best-effort traffic in congestion situations, i.e., when resources become scarce. Furthermore, LBE traffic gets a minimal share of bandwidth so that it will not be fully preempted by best-effort traffic. This is achieved by discarding LBE packets more aggressively than best-effort packets while trying to enqueue them in case of congestion.

There are numerous uses for this PHB, e.g., for transmission of background traffic such as bulk data transfers of minor

importance, backup data traffic during working hours or traffic caused by web search engines while gathering information from web servers. Thus, it constitutes the network equivalent to a background priority for processes in an operating system. Moreover, it is particularly useful in cases when packets of a better service are re-marked by a node

Bless & Wehrle

Expires: March 2000

[Page 1]

Internet-Draft

Lower Than Best-Effort PHB

September 1999

for subsequently receiving a forwarding treatment that is equivalent to the best-effort service. In this situation the LBE PHB helps to protect other best-effort packets from experiencing unfair forwarding treatment because it avoids their preemption by re-marked packets. For instance, this case occurs when heterogeneous multicast groups should be supported.

1 Purpose and Overview

In some situations changing a per-hop forwarding behavior (PHB) of an incoming packet is desired or required, so that the packet is subsequently forwarded with the lowest available priority. Usually, this forwarding behavior would be equivalent to the Default PHB resulting in a best-effort service. But, simply re-marking the DS codepoint (DSCP) to the DSCP value of the Default PHB will probably result in some unfair share of this re-marked traffic relating to best-effort traffic. This is due to the fact that nearly all packets which previously experienced a better service enter the behavior aggregate (BA) of the Default behavior. Consequently, other incoming packets carrying a DSCP of the Default PHB will be preempted by those re-marked packets if not enough resources are available for the combined traffic. This unfairness against existing best-effort traffic should be avoided.

The basic concept behind the proposed Lower-Than-Best-Effort per-hop behavior (LBE PHB) is to discard those re-marked packets more aggressively than packets belonging to the default PHB if resources become scarce. Merely discarding those packets more drastically in the re-marking node is not sufficient, because currently there may be enough resources available in this node, but maybe not in subsequent downstream nodes. Therefore, this re-marked traffic should be identifiable by a separate codepoint.

For instance, a re-marking of packets is required when a receiver joins a multicast group and does not want to or even is not able to receive

the currently used better service within this group (e.g., 1 Mbit/s Premium service). Instead of this, the receiver wants to get the traffic of this group without any quality of service guarantee, i.e., a best-effort service. The node which connects the new subtree of this receiver to the already existing multicast delivery tree must therefore degrade the quality of service of the incoming packet stream for replicated packets which are sent downstream on the output link of the newly joined subtree (see Fig. 1).

Moreover, specific excess traffic of any kind may be marked with the LBE codepoint. For example, this excess traffic could consist of packets belonging to non-responsive flows (e.g., UDP flows). Thus, other best-effort traffic (e.g., responsive TCP flows) is segregated from this excess traffic, consequently not being adversely affected by it.

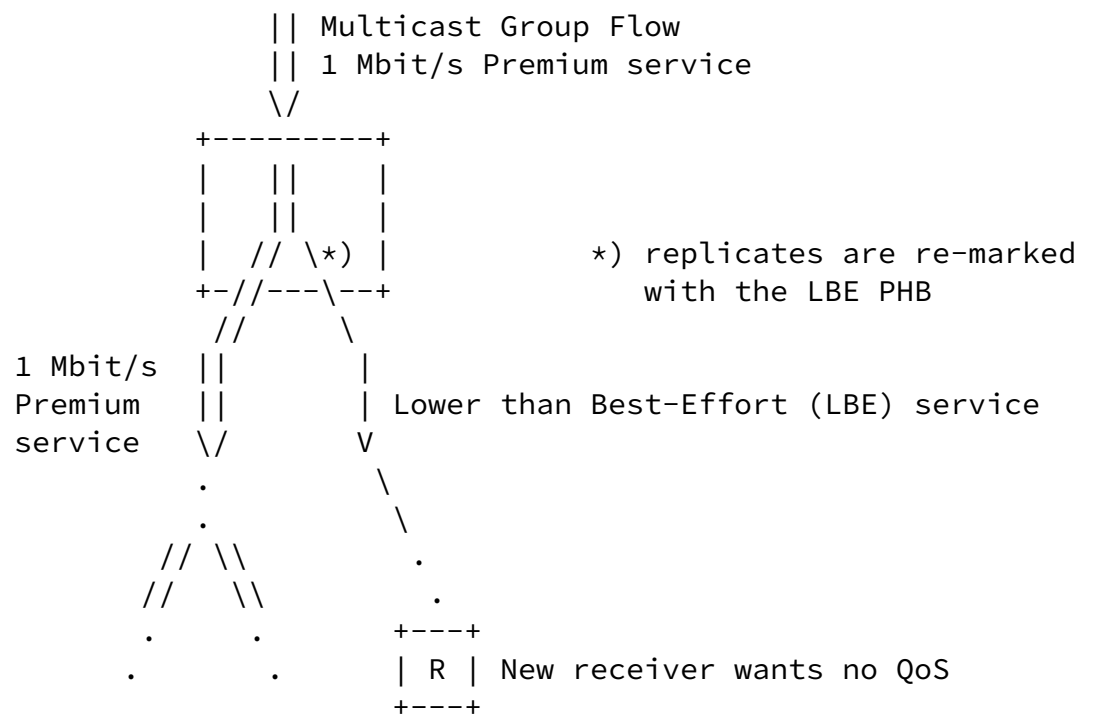


Figure 1: A new receiver without QoS requirements joining a multicast group that uses Premium service

As a further example, the LBE can be used as a separate service class for background traffic (e.g., bulk transfers of low priority) that should not impact the usual best-effort traffic. Examples for this kind of background traffic are backup data traffic sent over the network

during normal working hours and traffic from web search engines gathering data from web servers. Additionally, applications in DS aware end-systems are able to pre-mark packets of minor importance, while allowing to use the traditional inexpensive best-effort service for all other more important packets.

The LBE PHB solves the above mentioned problems by discarding LBE packets more rigorously than those of the best-effort BA in case of resource contention for residual bandwidth, therefore limiting the amount of packets in the LBE class in relation to the best-effort class. This is described more detailed in sections [2.1](#) and [2.3](#).

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 [\[4\]](#).

[2](#) Description of LBE per-hop behavior

A lower than best-effort service can be used in general to protect traffic in the best-effort service class from unfairness that would otherwise have been caused by those packets which are now in the LBE. The latter packets may stem from excess traffic or re-marked traffic that has been experiencing a better service before. Therefore, in order

to restrict the impact of those packets on packets in the Default PHB BA, the LBE PHB provides forwarding of IP packets with a higher drop precedence than Default PHB packets.

It is not necessary to have an LBE PHB group comprising more than one PHB, because within its BA, LBE PHB essentially resembles best-effort behavior that does not distinguish different types of traffic (e.g., responsive and non-responsive flows), too.

The Lower than Best-Effort per-hop behavior is intended for general use. There may be many cases in which the LBE PHB can be applied. Three of them were already mentioned in [section 1](#).

Using this PHB between two adjacent service providers is also useful, because the upstream domain possibly had enough resources left to carry the additional LBE traffic volume, whereas the downstream domain has not enough resources, therefore being in need of discarding some portion of this traffic.

2.1 Interaction with Other PHB Groups

The LBE PHB is essentially defined by its relation to the default PHB: an LBE packet is of minor relative importance compared to a best-effort packet. Thus, in case of contention for bandwidth (i.e., a congestion situation), packets receiving best-effort treatment preempt packets of the LBE behavior aggregate down to a certain share. This share must be considerably lower than that of the best-effort BA, but should not be equal to zero in order to retain a low portion of LBE traffic even when best-effort traffic takes up all available residual bandwidth.

Therefore, a minimum configured bandwidth share for LBE traffic exists as a lower bound that guarantees transport of LBE packets even in case of congestion. On the other hand, this bound also constitutes an upper limit for the share of LBE traffic during congestion. This upper bound may be static, that is, fixed in relation to the overall available bandwidth on a particular link (a constant value for $100-Y$ in Fig. 2), or it may be dynamic, i.e., fixed relative to the BE traffic share, thus variable in relation to the overall available bandwidth, because BE traffic may consume resources currently not used by other BAs (a dynamic value for Y that depends on the amount of BE traffic; corresponding to a constant ratio of $(100-Y)/(100-X)$ in Fig. 2).

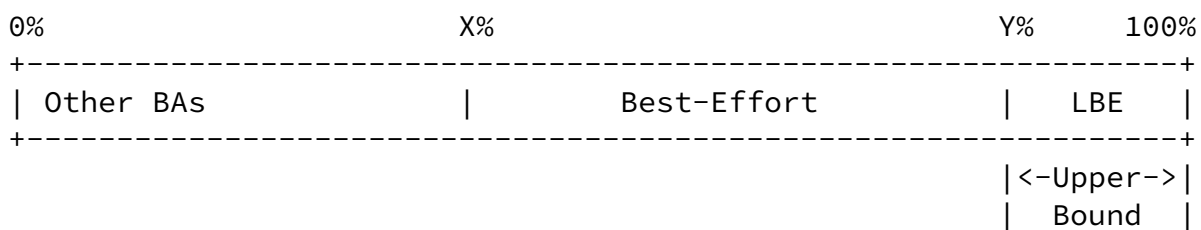
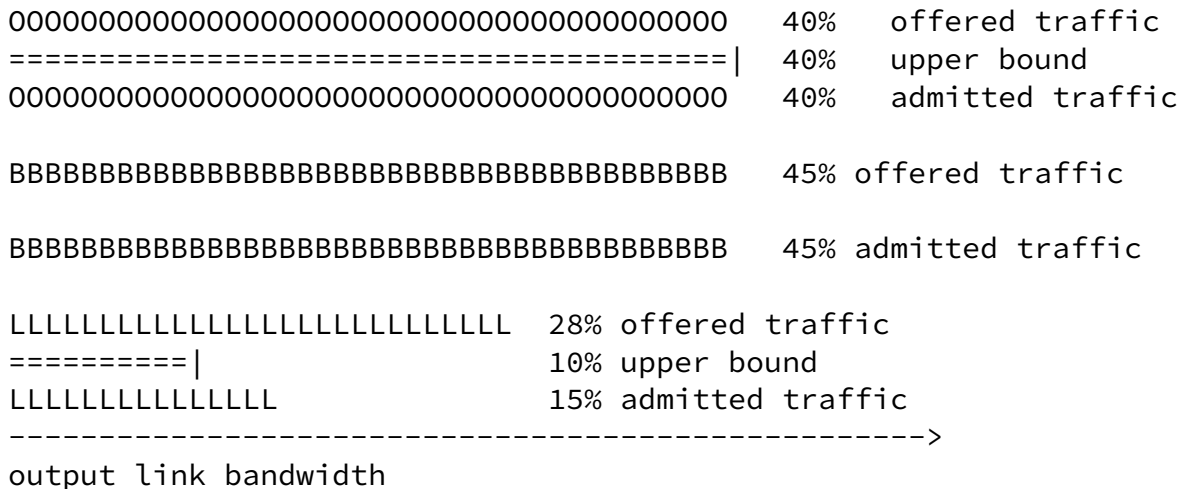


Figure 2: Bandwidth share limits for LBE traffic. Other behavior aggregates (BA) currently use $X\%$ of total available bandwidth, while best-effort and LBE traffic share the residual bandwidth.

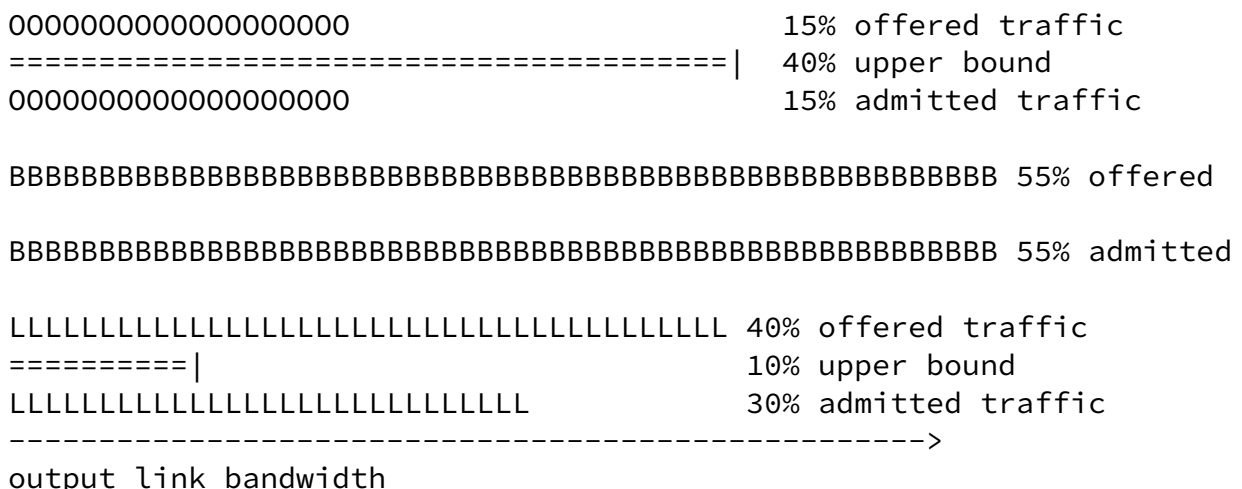
If enough resources are available for other BAs and BE traffic, i.e., residual bandwidth exists that is not used by best-effort traffic, LBE packets may use more than the previously defined bound on their share of total bandwidth, because this limit is only needed in case of congestion (see Fig. 3 and 4). Therefore, any remaining resources can be used to forward excess LBE traffic, because all BE demand is already met. In this case, there is no reason to discard any of the LBE packets until the residual bandwidth is exhausted by them, because their presence does

not adversely affect any of the best-effort packets.



O: Other BAs, B: Best-Effort BA, L: LBE BA

Figure 3: Other BAs fully exploit their allotted resources, LBE BA gets residual bandwidth that best-effort doesn't use.



O: Other BAs, B: Best-Effort BA, L: LBE BA

Figure 4: Other BAs do not fully use their allotted resources, best-effort utilizes unused bandwidth of the other BAs, LBE gets residual bandwidth

The primary objective of the LBE PHB is to segregate packets of the best-effort behavior aggregate from packets which should receive a lower than best-effort treatment in case of congestion. This is achieved by

using a higher drop precedence for LBE packets than for best-effort packets, and by limiting the overall amount of LBE traffic in relation to the best-effort behavior aggregate (see Fig. 5). Therefore, a congested node tries to protect best-effort packets from being lost by preferably discarding LBE packets.

[illegible][illegible][illegible][illegible]

output link bandwidth

0: Other BAs, B: Best-Effort BA, L: LBE BA

Figure 5: Other BAs do not fully use their allotted resources, best-effort utilizes unused bandwidth of the other BAs, LBE is limited to its specified upper bound.

With exception of the Default PHB, the LBE PHB is entirely independent of all other existing PHB specifications. Thus, any other PHB groups may co-exist with the LBE PHB in the same DS domain, because the LBE PHB does not preempt forwarding resources of other PHB groups.

If only a part of all packets belonging to a microflow are marked as LBE, the probability for reordering this microflow's packets may be increased in dependence on the relation of the prior PHB to the LBE PHB and may depend on the actual implementation. However, because the LBE PHB is defined by a special relation to the Default PHB, it is recommended that packets of a microflow are not reordered if they are marked by codepoints associated with these two PHBs.

A realization of the LBE PHB may use an already existing PHB of higher drop precedence (e.g., AFx2 or AFx3 from the AF PHB group [2]), if its actual implementation fulfills the requirements of this specification.

2.2 Traffic Conditioning Actions

Usually, the amount of LBE traffic is implicitly limited by queueing mechanisms and related discard actions. Therefore, there is normally no need to meter and police LBE traffic explicitly. However, a DS domain MAY control the amount of LBE traffic that enters or exits the domain. Traffic conditioning actions MAY include traffic shaping and discarding of packets.

[2.3](#) Queueing and Discard Behavior

All packets of an arbitrary microflow that are marked for the LBE PHB MUST NOT be reordered. The dropping algorithm MUST treat all packets within the LBE BA identically, i.e., the discard rate of a particular microflow's packets will be proportional to that flow's percentage of the total amount of traffic with this LBE BA.

It is recommended that LBE packets use the same queue as best-effort packets in order to avoid reordering of a microflow's packets that are marked intermittently for best-effort or LBE forwarding.

One way to achieve the above objectives is to use a common queue for best-effort and LBE packets that is actively managed by a Random Early Detection (RED) scheme [5]. A separate RED parameter set is managed for each traffic type. If the "current" queue length (usually a weighted average value) grows beyond a lower threshold, new arriving LBE packets for this queue are going to be dropped randomly with increasing probability, until the queue length reaches an upper threshold. Beyond this upper threshold every new incoming LBE packet for this queue is going to be discarded. If the threshold values for LBE packets are lower than those for best-effort packets, the RED queue will start discarding packets earlier.

[2.4](#) Recommended Codepoint for this PHB

As long as this a PHB proposal, the temporary recommended code point is taken from the experimental/local use (EXP/LU) portion of the codepoint space. The recommended temporary codepoint is '101011' (see section 3 of [1] for the meaning of this notation).

[2.5](#) Configuration and Management Issues

There is no need for providing any resource-based admission control mechanisms for this PHB. As described in [section 2.1](#), a configurable minimal bandwidth share, respectively upper bound, exists for bandwidth used by the LBE BA if no residual bandwidth is left and all other bandwidth is used for best-effort.

[3](#) Security Considerations

Basically, the LBE PHB causes no other security implications besides the ones already mentioned in [\[1\]](#). Because LBE PHB provides a quality that is even lower compared to the usual best-effort delivery, there is now one more possible PHB to reduce a packet's service.

Bless & Wehrle

Expires: March 2000

[Page 7]

Internet-Draft

Lower Than Best-Effort PHB

September 1999

On the other hand, there is currently no PHB providing a worse service, so LBE packets cannot be further reduced in service by re-marking. Consequently, re-marking the inner header's codepoint of LBE packets at the egress of a tunnel with a codepoint of another PHB would have only a positive effect on these packets. However, this would possibly eliminate the primary objective of the LBE and lead to depletion of forwarding resources for other traffic streams in congestion situations.

[4](#) Tunneling

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

[5](#) Implications on Services

As described in [section 2.1](#), traffic using the current best-effort service should be segregated and protected from LBE traffic in congestion situations. Therefore, performance of the common best-effort service is increased under those conditions.

[6](#) Mapping to link-layer QoS mechanisms

In a shared medium LBE traffic has a very good counterpart in the link-layer QoS mechanisms as defined by IEEE 802.1p: the background traffic type. Therefore, packets carrying a DSCP value of the LBE PHB can be mapped to 802.1p background traffic priority, i.e., setting the "user_priority" field to value 1 in all link-layer frames that carry a part of an LBE packet as payload. In order to achieve a real support for LBE packets, link-layer frames containing best-effort packets should use the default user_priority of 0 for indicating traffic type "Best Effort".

LBE packets within a switched link-layer could also use available means to indicate a higher drop precedence for LBE packets. For instance, when using ATM as link-layer, the value encoded in the Cell Loss Priority (CLP) field of the ATM cell header [6] may be set to 1 in all ATM cells carrying a part of an LBE packet as payload, whereas cells carrying a part of a best-effort packet as payload should use a CLP value of 0.

As another example, when using Frame Relay as link-layer, the Discard Eligibility Indicator (DE) bit can be set to 1 for frames containing an LBE packet as payload, indicating that this frame should be discarded in preference to other frames in a congestion situation [7]. All frames carrying a part of a best-effort packet as payload should use a DE bit value of 0.

[7](#) Acknowledgments

Thanks to Jochen Schiller for discussion of the LBE PHB.

References

- [1] F. Baker, D. Black, S. Blake, and K. Nichols. Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. [RFC 2474](#), Dec. 1998.
- [2] F. Baker, J. Heinanen, W. Weiss, and J. Wroclawski. Assured Forwarding PHB Group. [RFC2597](#), June 1999.
- [3] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Services. [RFC 2475](#), Dec. 1998.

- [4] S. Bradner. Key words for use in RFCs to Indicate Requirement Levels. [RFC 2119](#), Mar. 1997.
- [5] S. Floyd and V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. IEEE/ACM Transaction on Networking, 1(4):397--413, Aug. 1993.
- [6] ITU-T. B-ISDN ATM Layer Specification. ITU-T Recommendation I.361, 11/1995.
- [7] ITU-T. ISDN Data Link Layer Specification for Frame Mode Bearer Services. ITU-T Recommendation Q.922, 1992.

[8](#) Authors' addresses

Comments and questions related to this draft can be addressed to one of the authors listed below.

Roland Bless
Institute of Telematics
Universitaet Karlsruhe (TH)
Zirkel 2
D-76128 Karlsruhe
Germany
Tel.: +49 721 608 6396
bless@telematik.informatik.uni-karlsruhe.de

Klaus Wehrle
Institute of Telematics
Universitaet Karlsruhe (TH)
Zirkel 2
D-76128 Karlsruhe
Germany
Tel.: +49 721 608 6414
wehrle@telematik.informatik.uni-karlsruhe.de