

Transport Area Working Group
Internet-Draft
Intended status: Informational
Expires: January 31, 2021

G. White, Ed.
CableLabs
July 30, 2020

Operational Guidance for Deployment of L4S in the Internet
draft-white-tsvwg-l4sops-00

Abstract

This is an early, work-in-progress draft - a start at getting some of the ideas from the mailing list and email exchanges on paper.

This draft is intended to provide guidance to operators of end-systems, operators of networks, and researchers in order to ensure reasonable fairness between L4S and Classic flows sharing a single-queue [RFC3168](#) bottleneck link. This draft identifies opportunities to prevent and/or detect and resolve fairness problems in such networks.

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 <https://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 January 31, 2021.

Copyright Notice

Copyright (c) 2020 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 (<https://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	2
2.	Per-Flow Fairness	3
3.	Operator of an L4S host	3
3.1.	CDN Servers	3
3.2.	Other hosts	4
4.	Operator of a Network	4
4.1.	Configure AQM to treat ECT1 as NotECT	4
4.2.	Configure Non-Coupled Dual Queue	5
4.3.	WRED with ECT1 Differentiation	5
4.4.	ECT1 Tunnel Bypass	6
4.5.	Disable RFC3168 ECN Marking	6
4.6.	Re-mark ECT1 to NotECT Prior to AQM	6
5.	Researchers	6
5.1.	Detection of Classic ECN FIFO Bottlenecks	6
5.2.	End-to-end measurement of L4S vs. Classic performance . .	6
6.	Contributors	6
7.	IANA Considerations	6
8.	Security Considerations	7
9.	Informative References	7
	Author's Address	7

[1.](#) Introduction

In the majority of network paths, including paths where the bottleneck link utilizes packet drops (either due to buffer overrun or active queue management) in response to congestion, as well as paths that implement a 'flow-queuing' scheduler such as fq_codel or Cobalt, and those that implement dual-Q-coupled AQM, L4S traffic coexists well with classic congestion controlled traffic.

On network paths where the bottleneck link implements a shared-queue (FIFO) with an Active Queue Management algorithm that provides Explicit Congestion Notification signaling according to [RFC3168](#), it has been demonstrated that when a set of long-running flows comprising both "Classic" congestion controlled flows and L4S-compliant congestion controlled flows compete for bandwidth, the classic congestion controlled flows may achieve lower throughput when compared to the L4S congestion controlled flows. This 'unfairness' between the two classes appears to be more pronounced on longer RTT paths (e.g. 50ms and above) and/or at higher link rates (e.g. 50 Mbps and above).

White

Expires January 31, 2021

[Page 2]

The root cause of this unfairness is that [RFC3168](#) does not differentiate between packets marked ECT0 (used by classic senders) and those marked ECT1 (used by L4S senders), and provides an identical congestion signal (CE marks) to both classes, whereas the two classes respond differently to that congestion signal. The classic senders expect that CE marks are sent very rarely (e.g. approximately 1 CE mark every 200 round trips on a 50 Mbps x 50ms path) while the L4S senders expect very frequent CE marking (e.g. approximately 2 CE marks per round trip). The result is that the classic senders respond to the CE marks provided by the bottleneck by yielding capacity to the L4S flows. While this has not been demonstrated to cause starvation of the classic flows, the resulting rate imbalance can be a cause of concern.

2. Per-Flow Fairness

There are a number of factors that influence the relative rates achieved by a set of congestion controlled flows sharing a queue in a bottleneck link.

TODO: discuss startup & convergence times, short flows, RTT-unfairness, differences in deployed CC algorithms, etc.

TODO: also mention that flow sharding is commonplace, so per-flow fairness does not imply per-application fairness

3. Operator of an L4S host

Support for L4S involves both endpoints: ECT1 marking & L4S-compatible congestion control on the sender, and ECN feedback on the receiver. Between these two entities, it is incumbent upon the sender to evaluate the potential for unfairness and make decisions whether or not to use L4S congestion control. The receiver is not expected to perform any testing or monitoring for unfairness, and is also not expected to invoke any active response in the case that unfairness occurs.

3.1. CDN Servers

Some hosts (such as CDN leaf nodes and servers internal to an ISP) are deployed in environments in which they serve content to a constrained set of networks or clients. The operator of such hosts may be able to determine whether there is the possibility of [RFC3168](#) FIFO bottlenecks being present, and utilize this information to make decisions on selectively deploying L4S.

- o Prior to deploying L4S on servers:

White

Expires January 31, 2021

[Page 3]

- * Consult with network operators on presence of [RFC3168](#) FIFO bottlenecks
- * Perform downstream tests per access network
 - + Tests (TBD) to detect absence of [RFC 3168](#)
 - + Enable AccECN feedback, but enable/disable L4S per access network
- o In-band [RFC3168](#) detection and monitoring:
 - * Real-time response (fallback)
 - * Non-real-time response (disable for future connections)

3.2. Other hosts

- o In-band [RFC3168](#) detection (and possibly fallback)
- o Per-dst path test:
 - * For a connection capable of L4S feedback
 - * If CE feedback, perform active test (TBD) for [RFC3168](#) presence
 - * Could cache result per-dst
- o Query a TBD public whitelist of domains that are participating in L4S experiment

4. Operator of a Network

While it is, of course, preferred for networks to deploy L4S-capable high fidelity congestion signaling, a network operator who has deployed equipment in a likely bottleneck link location (i.e. a link that is expected to be fully saturated) that is configured with an [RFC3168](#) FIFO AQM can take certain steps in order to improve rate fairness between classic traffic and L4S traffic.

4.1. Configure AQM to treat ECT1 as NotECT

If equipment is configurable in such a way as to only supply CE marks to ECT0 packets, and treat ECT1 packets identically to NotECT, or is upgradable to support this capability, doing so will eliminate the risk of unfairness.

4.2. Configure Non-Coupled Dual Queue

Equipment supporting [RFC3168](#) may be configurable to enable two parallel queues for the same traffic class, with classification done based on the ECN field.

Option 1:

- o Configure 2 queues, both with ECN; 50:50 WRR scheduler
- o Queue #1: ECT1 & CE packets - Shallow immediate AQM target
- o Queue #2: ECT0 & NotECT packets - Classic AQM target
- o Outcome
 - * n L4S flows and m long-running Classic flows
 - * if m & n are non-zero, get $1/2n$ and $1/2m$ of the capacity, otherwise $1/n$ or $1/m$
 - * never $< 1/2$ each flow's rate if all had been Classic

Option 2:

- o Configure 2 queues, both with AQM; 50:50 WRR scheduler
- o Queue #1: ECT1 & NotECT packets - ECN disabled
- o Queue #2: ECT0 & CE packets - ECN enabled
- o Outcome
 - * ECT1 treated as NotECT
 - * Flow balance for the 2 queues the same as in option 1

4.3. WRED with ECT1 Differentiation

This configuration is similar to Option 2 in the previous section, but uses a single queue with WRED functionality.

- o Configure the queue with two WRED classes
- o Class #1: ECT1 & NotECT packets - ECN disabled
- o Class #2: ECT0 & CE packets - ECN enabled

4.4. ECT1 Tunnel Bypass

Using an [RFC6040](#) compatibility mode tunnel, tunnel ECT1 traffic through the [RFC3168](#) bottleneck with the outer header indicating Not-ECT.

Two variants

1. per-domain: tunnel ECT1 pkts to domain edge towards dst
2. per-dst: tunnel ECT1 pkts to dst

4.5. Disable [RFC3168](#) ECN Marking

While not a recommended alternative, disabling [RFC3168](#) ECN marking eliminates the fairness issue. Clearly a downside to this approach is that classic senders will no longer get the benefits of Explicit Congestion Notification.

4.6. Re-mark ECT1 to NotECT Prior to AQM

While not a recommended alternative, remarking ECT1 packets as NotECT ensures that they are treated identically to classic NotECT senders. However, this also eliminates the possibility of downstream L4S bottlenecks providing high fidelity congestion signals.

5. Researchers

5.1. Detection of Classic ECN FIFO Bottlenecks

TODO: Describe active testing methods, in-band or out-of-band, that can distinguish FIFO from FQ.

5.2. End-to-end measurement of L4S vs. Classic performance

TBD

6. Contributors

Thanks to Bob Briscoe, Jake Holland, Koen De Schepper, Olivier Tilmans, Tom Henderson, Asad Ahmed, and members of the TSVWG mailing list for their contributions to this document.

7. IANA Considerations

None.

8. Security Considerations

None.

9. Informative References

[I-D.ietf-tsvwg-aqm-dualq-coupled]

Schepper, K., Briscoe, B., and G. White, "DualQ Coupled AQMs for Low Latency, Low Loss and Scalable Throughput (L4S)", [draft-ietf-tsvwg-aqm-dualq-coupled-12](#) (work in progress), July 2020.

[I-D.ietf-tsvwg-ecn-l4s-id]

Schepper, K. and B. Briscoe, "Identifying Modified Explicit Congestion Notification (ECN) Semantics for Ultra-Low Queuing Delay (L4S)", [draft-ietf-tsvwg-ecn-l4s-id-10](#) (work in progress), March 2020.

[I-D.ietf-tsvwg-l4s-arch]

Briscoe, B., Schepper, K., Bagnulo, M., and G. White, "Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service: Architecture", [draft-ietf-tsvwg-l4s-arch-06](#) (work in progress), March 2020.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in [RFC 2119](#) Key Words", [BCP 14](#), [RFC 8174](#), DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

[RFC8290] Hoeiland-Joergensen, T., McKenney, P., Taht, D., Gettys, J., and E. Dumazet, "The Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm", [RFC 8290](#), DOI 10.17487/RFC8290, January 2018, <<https://www.rfc-editor.org/info/rfc8290>>.

Author's Address

Greg White (editor)
CableLabs

Email: g.white@cablelabs.com

White

Expires January 31, 2021

[Page 7]