

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 27, 2015

M. Welzl
University of Oslo
G. Fairhurst
University of Aberdeen
October 24, 2014

**The Benefits and Pitfalls of using Explicit Congestion Notification
(ECN)
draft-ietf-aqm-ecn-benefits-00**

Abstract

This document describes the potential benefits and pitfalls when applications enable Explicit Congestion Notification (ECN). It outlines the principal gains in terms of increased throughput, reduced delay and other benefits when ECN is used over network paths that include equipment that supports ECN-marking. It also lists potential problems that might occur when ECN is used. The document does not propose new algorithms that may be able to use ECN or describe the details of implementation of ECN in endpoint devices, routers and other network devices.

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 April 27, 2015.

Copyright Notice

Copyright (c) 2014 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.

1. Introduction

Internet Transports (such as TCP and SCTP) have two ways to detect congestion: the loss of a packet and, if Explicit Congestion Notification (ECN) [[RFC3168](#)] is enabled, by reception of a packet with a Congestion Experienced (CE)-marking in the IP header. Both of these are treated by transports as indications of (potential) congestion. ECN may also be enabled by other transports. UDP applications may enable ECN when they are able to correctly process the ECN signals (e.g. ECN with RTP [[RFC6679](#)]).

When an application enables the use of ECN, the transport layer sets the ECT(0) or ECT(1) codepoint in the IP header of packets that it sends to indicate to routers that they may mark, rather than drop, packets in periods of congestion. This marking is generally performed by Active Queue Management (AQM) [[RFC2309.bis](#)] and may be the result of various AQM algorithms, where the exact combination of AQM/ECN algorithms is generally not known by the transport endpoints.

ECN makes it possible for the network to signal congestion without packet loss. This lets the network deliver some packets to an application that would otherwise have been dropped. This packet loss reduction is the most obvious benefit of ECN, but it is often relatively modest. However, enabling ECN can also result in a number of beneficial side-effects, some of which may be much more significant than the immediate packet loss reduction from ECN-marking instead of dropping packets. Several of these benefits have to do with reducing latency in some way (e.g. reduced Head-of-Line Blocking and potentially smaller queuing delay, depending on the marking rules in routers). There are also some potential pitfalls when enabling ECN.

The focus of this document is on usage of ECN, not its implementation in endpoint devices, routers and other network devices. [[RFC3168](#)] describes a method in which a router sets the CE codepoint of an ECN-Capable packet at the time that the router would otherwise have dropped the packet.

While it has often been assumed that routers mark packets at the same level of congestion at which they would otherwise drop them, separate

configuration of the drop and mark thresholds is known to be supported in some network devices and this is recommended in [\[RFC2309.bis\]](#). Some benefits of ECN that are discussed rely upon routers marking packets at a lower level of congestion before they would otherwise drop packets from queue overflow [\[KH13\]](#).

Some of benefits are also only realised when the transport endpoint behaviour is also updated, this is discussed further in [Section 5](#).

The remainder of this document discusses the potential for ECN to positively benefit an application without making specific assumptions about configuration or implementation.

[2.](#) ECN Deployment Scenarios / Use Cases

XXX to be continued -- this section is intended to describe some specific example cases of where ECN has provided benefit XXX

[2.1.](#) ECN within data centers

ECN with a low marking threshold has been proposed for use within a data centre environment. This proposed usage exploits ECN in combination with an updated transport behaviour, Datacenter TCP (DCTCP) [\[AL10\]](#).

[3.](#) Benefit of using ECN to avoid congestion loss

An application that uses a transport that supports ECN can benefit in several ways:

[3.1.](#) Improved Throughput

ECN can improve the throughput performance of applications, although this increase in throughput offered by ECN is often not the most significant gain.

When an application uses a light to moderately loaded network path, the number of packets that are dropped due to congestion is small. Using an example from Table 1 of [\[RFC3649\]](#), for a standard TCP sender with a Round Trip Time, RTT, of 0.1 seconds, a packet size of 1500 bytes and an average throughput of 1 Mbps, the average packet drop ratio is 0.02. This translates into an approximate 2% throughput gain if ECN is enabled. In heavy congestion, packet loss may be unavoidable with, or without, ECN.

3.2. Reduced Head-of-Line Blocking

Many transports provide in-order delivery of received data to the applications they support. This requires that the transport stalls (or waits) for all data that was sent ahead of a particular segment to be correctly received before it can forward any later data. This is the usual requirement for TCP and SCTP. PR-SCTP [[RFC3758](#)], UDP, and DCCP [[RFC4340](#)] provide a transport that does not have this requirement.

Delaying data to provide in-order transmission to an application results in latency when segments are dropped as indications of congestion. The congestive loss creates a delay of at least one RTT for a loss event before data can be delivered to an application. We call this Head-of-Line (HOL) blocking.

In contrast, using ECN can remove the resulting delay for a loss that is a result of congestion:

- o First, the application receives the data normally - this also avoids dropping data that has already made it across at least part of the network path. This avoids the additional delay of waiting for recovery of the lost segment.
- o Second, the transport receiver notes the ECN-marked packets, and then requests the sender to make an appropriate congestion-response for future traffic.

3.3. Reduced Probability of RTO Expiry

ECN can help reduce the chance of the TCP or SCTP retransmission timer expiring (RTO expiry). When an application sends a burst of segments and then becomes idle (either because the application has no further data to send or the network prevents sending further data - e.g. flow or congestion control at the transport layer), the last segment of the burst may be lost. It is often not possible to recover the last segment (or last few segments) using standard methods such as Fast Recovery, since the receiver is unaware that the lost segments were actually sent.

ECN provides a mitigation when the loss is a result of (mild) congestion, since a router may mark, rather than drop, these segments - which benefits the application in a way similar to above, but with the significant additional benefit that this eliminates a retransmission event. The application benefits because:

- o Data is received without HOL blocking.
- o The transport does not suffer RTO expiry with consequent loss of state about the network path it is using. This would cause it to reset path estimates such as the RTT, the congestion window, and possibly other transport state that can reduce the performance of the transport until it adapts to the path again. This can improve the throughput of the application.

The benefit of avoiding reliance on an RTO-based retransmission event can be especially significant when ECN is used on TCP SYN/ACK packets as specified in [[RFC5562](#)] because in this case TCP cannot base its RTO for these packets on prior RTT measurements from the same connection.

3.4. Applications that do not retransmit lost packets

Certain latency-critical applications do not retransmit lost packets, yet they may be able to adjust the sending rate in the presence of congestion. Examples of such applications include UDP-based services that carry Voice over IP (VoIP), interactive video or real-time data. By decoupling congestion control from loss, ECN can allow such applications to reduce their rate before experiencing significant loss. It also enables them to decide how to discard data in a controlled manner, rather than forcing them to recover from loss. This reduces the negative impact of having to rely on loss-hiding mechanisms (e.g. Packet forward error correction, or data duplication), yielding a direct positive impact on the quality experienced by the users of these applications.

4. Benefit from Early Congestion Detection

If ECN is configured such that routers mark packets at a lower level of congestion before they would otherwise drop packets from queue overflow, an application can benefit from using ECN in the following ways:

4.1. Avoiding Capacity Overshoot

ECN can help capacity probing algorithms (such as Slow Start) from significantly exceeding the bottleneck capacity of a network path. Since a transport that enables ECN can receive congestion signals before there is serious congestion, an early-marking method can help a transport respond before it induces significant congestion. For example, a TCP or SCTP sender can avoid incurring significant congestion during Slow Start, or a bulk application that tries to increase its rate as fast as possible, may detect the presence of

congestion, causing it to reduce its rate.

Use of ECN is more effective than schemes such as Limited Slow-Start [[RFC3742](#)] because it provides direct information about the state of the network path. An ECN-enabled application probing for bandwidth can reduce its rate as soon as ECN-marked packets are detected, and before the applications increases its rate to the point where it builds a router queue that induces congestion loss. This benefits the application seeking to increase its rate - but perhaps more significantly, it eliminates the often unwanted loss and queueing delay that otherwise may be inflicted on flows that share a common bottleneck.

4.2. Making Congestion Visible

A characteristic of using ECN is that it exposes the presence of congestion on a network path to the transport and network layers. This information could be used for monitoring performance of the path, and could be used to directly meter the amount of congestion that has been encountered upstream on a path; metering packet loss is harder. This is used by Congestion Exposure (CoNex) [[RFC6789](#)].

Note: traffic that observes only congestion marks and no loss implies that a sender is experiencing only congestion and not other sources of packet loss (e.g. link corruption or loss in middleboxes). The converse is not true - a mixture of ECN-marks and loss may occur during only congestion or from a combination of packet loss and congestion.

5. Other forms of ECN-Marking/Reactions

The ECN mechanism defines both how packets are marked and transports need to react to markings. This section describes the benefits when updated methods are used.

Benefit has been noted when packets are marked earlier than they would otherwise be dropped, using an instantaneous queue, and if the receiver provides precise feedback about the number of packet marks encountered, a better sender behavior is possible. This has been shown by Datacenter TCP (DCTCP) [[AL10](#)].

Precise feedback about the number of packet marks encountered is supported by RTP over UDP [[RFC6679](#)] and proposed for SCTP [[ST14](#)] and TCP [[KU13](#)]. An underlying assumption of DCTCP is that it is deployed in confined environments such as a datacenter. It is currently unknown whether or how such behaviour could be safely introduced into the Internet.

6. Pitfalls when using ECN

This section describes issues with ECN.

6.1. Bleaching and middlebox requirements to deploy

Cases have been noted where a sending endpoint marks a packet with a non-zero ECN mark, but the packet is received with a zero ECN value by the remote endpoint.

The current IPv4 and IPv6 specifications assign usage of 2 bits in the IP header to carry the ECN codepoint. A previous usage assigned these bits as a part of the now deprecated Type of Service (ToS) field. Equipment conformant with this older specification may remark or erase the ECN codepoints, such equipment needs to be updated to the current specifications to support ECN.

Another policy may erase or "bleach" the ECN marks at a network edge (resetting these to zero) for various reasons (including normalising packets to hide which equipment support ECN). This policy prevents use of ECN.

Some networks may use ECN internally or tunnel ECN for traffic engineering or security. Guidance on the correct use of ECN in this case is provided in [[RFC6040](#)].

The recommendation of this document is that a router or middlebox **MUST** not change a packet with a CE mark to a zero codepoint (if the CE marking is not propagated, the packet **MUST** be discarded), and **SHOULD NOT** remark an ECT(0) or ECT(1) mark to zero.

6.2. Cheating

Endpoint receivers **MUST NOT** try to conceal reception of CE marks in the ECN feedback information they provide to the sending endpoint. Transport protocols are actively encouraged to include mechanisms that can detect and appropriately respond to such misbehavior.

6.3. The possible need to verify if a path really supports ECN

Endpoints need to be robust to path changes that may impact the ability to effectively signal or use ECN across a path, e.g. when a path changes to use a middlebox that bleaches ECN codepoints. As a necessary but short term fix, such mechanisms could fall-back to disabling use of ECN.

7. Conclusion

People configuring host stacks and network devices should ensure that their equipment correctly reacts to packets carrying ECN codepoints. This includes:

- o routers not resetting the ECN codepoint to zero by default
- o routers correctly updating the codepoint in the presence of congestion
- o routers correctly supporting alternate ECN semantics ([[RFC4774](#)])
- o hosts receiving ECN marks correctly reflecting them

Application developers should where possible use transports that enable the benefits of ECN. Once enabled, the benefits of ECN are provided by the transport layer and the application does not need to be rewritten to gain these benefits. Table 1 summarises some of these benefits.

+-----+-----+-----+-----+-----+				
Section	Benefit			
+-----+-----+-----+-----+-----+				
2.1	Improved Throughput			
2.2	Reduced Head-of-Line			
2.3	Reduced Probability of RTO Expiry			
2.4	Applications that do not retransmit lost packets			
3.1	Avoiding Capacity Overshoot			
3.2	Making Congestion Visible			
+-----+-----+-----+-----+-----+				

Table 1: Summary of Key Benefits

8. Acknowledgements

The authors were part-funded by the European Community under its Seventh Framework Programme through the Reducing Internet Transport Latency (RITE) project (ICT-317700). The views expressed are solely those of the authors.

The authors would like to thank the following people for their comments on prior versions of this document: Bob Briscoe, David Collier-Brown, John Leslie, Colin Perkins, Richard Scheffenegger, Dave Taht

9. IANA Considerations

XXRFC ED - PLEASE REMOVE THIS SECTION XXX

This memo includes no request to IANA.

10. Security Considerations

This document introduces no new security considerations. Each RFC listed in this document discusses the security considerations of the specification it contains.

11. References

11.1. Normative References

[RFC2309.bis]

Baker, F. and G. Fairhurst, "IETF Recommendations Regarding Active Queue Management", [draft-ietf-aqm-recommendation-06](#) (work in progress), October 2014.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", [RFC 3168](#), September 2001.

11.2. Informative References

[AL10] Alizadeh, M., Greenberg, A., Maltz, D., Padhye, J., Patel, P., Prabhakar, B., Sengupta, S., and M. Sridharan, "Data Center TCP (DCTCP)", SIGCOMM 2010, August 2010.

[KH13] Khademi, N., Ros, D., and M. Welzl, "The New AQM Kids on the Block: Much Ado About Nothing?", University of Oslo Department of Informatics technical report 434, October 2013.

[KU13] Kuehlewind, M. and R. Scheffenegger, "Problem Statement and Requirements for a More Accurate ECN Feedback", [draft-ietf-tcpm-accecn-reqs-04.txt](#) (work in progress), October 2013.

[RFC3649] Floyd, S., "HighSpeed TCP for Large Congestion Windows", [RFC 3649](#), December 2003.

[RFC3742] Floyd, S., "Limited Slow-Start for TCP with Large

Congestion Windows", [RFC 3742](#), March 2004.

- [RFC3758] Stewart, R., Ramalho, M., Xie, Q., Tuexen, M., and P. Conrad, "Stream Control Transmission Protocol (SCTP) Partial Reliability Extension", [RFC 3758](#), May 2004.
- [RFC4340] Kohler, E., Handley, M., and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", [RFC 4340](#), March 2006.
- [RFC4774] Floyd, S., "Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field", [BCP 124](#), [RFC 4774](#), November 2006.
- [RFC5562] Kuzmanovic, A., Mondal, A., Floyd, S., and K. Ramakrishnan, "Adding Explicit Congestion Notification (ECN) Capability to TCP's SYN/ACK Packets", [RFC 5562](#), June 2009.
- [RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", [RFC 6040](#), November 2010.
- [RFC6679] Westerlund, M., Johansson, I., Perkins, C., O'Hanlon, P., and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", [RFC 6679](#), August 2012.
- [RFC6789] Briscoe, B., Woundy, R., and A. Cooper, "Congestion Exposure (ConEx) Concepts and Use Cases", [RFC 6789](#), December 2012.
- [ST14] Stewart, R., Tuexen, M., and X. Dong, "ECN for Stream Control Transmission Protocol (SCTP)", [draft-stewart-tsvwg-sctpecn-05.txt](#) (work in progress), January 2014.

Authors' Addresses

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

Phone: +47 22 85 24 20
Email: michawe@ifi.uio.no

Godred Fairhurst
University of Aberdeen
School of Engineering, Fraser Noble Building
Aberdeen, AB24 3UE
UK

Phone:

Email: gorry@erg.abdn.ac.uk