Transport Area Working Group
Internet-Draft
Obsoletes: 3540 (if approved)
Updates: 3168, 4341, 4342, 5622, 6679
(if approved)
Intended status: Standards Track
Expires: September 9, 2017

D. Black
Dell EMC

March 8, 2017

Explicit Congestion Notification (ECN) Experimentation
draft-ietf-tsvwg-ecn-experimentation-01

Abstract

Multiple protocol experiments have been proposed that involve changes to Explicit Congestion Notification (ECN) as specified in RFC 3168. This memo summarizes the proposed areas of experimentation to provide an overview to the Internet community and updates RFC 3168, a Proposed Standard RFC, to allow the experiments to proceed without requiring a standards process exception for each Experimental RFC to update RFC 3168. Each experiment is still required to be documented in an Experimental RFC. In addition, this memo makes related updates to the ECN specifications for RTP in RFC 6679 and to the ECN specifications for DCCP in RFC 4341, RFC 4342 and RFC 5622. This memo also records the conclusion of the ECN Nonce experiment in RFC 3540, obsoletes RFC 3540 and reclassifies it as Historic to enable new experimental use of the ECT(1) codepoint.

Table of Contents

1. Introduction

Multiple protocol experiments have been proposed that involve changes to Explicit Congestion Notification (ECN) as specified in RFC 3168 [RFC3168]. This memo summarizes the proposed areas of experimentation to provide an overview to the Internet community and updates RFC 3168 to allow the experiments to proceed without requiring a standards process exception for each Experimental RFC to update RFC 3168, a Proposed Standard RFC. This memo also makes related updates to the ECN specification for RTP in RFC 6679 [RFC6679] for the same reason. Each experiment is still required to be documented in one or more separate RFCs, but use of Experimental RFCs for this purpose does not require a process exception to modify RFC 3168 or RFC 6679 when the modification falls within the bounds established by this memo.

One of these areas of experimentation involves use of the ECT(1) codepoint that was dedicated to the ECN Nonce experiment as described in RFC 3540 [RFC3540]. This memo records the conclusion of the ECN Nonce experiment, obsoletes RFC 3540 and reclassifies it as Historic.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Scope of ECN Experiments

Three areas of ECN experimentation are covered by this memo; the cited Internet-Drafts should be consulted for the goals and rationale of each proposed experiment:

Congestion Response Differences: For congestion indicated by ECN, use a different IETF-approved sender congestion response (e.g., reduce the response so that the sender backs off by a smaller amount) by comparison to congestion indicated by loss, e.g., as proposed in [I-D.khademi-tcpm-alternativebackoff-ecn] and [I-D.briscoe-tsvwg-ecn-l4s-id] - the experiment in the latter draft couples the backoff change to ECT Differences functionality (next bullet). This is at variance with RFC 3168's requirement that a sender's congestion control response to ECN congestion indications be the same as to drops.

ECT Differences: Use ECT(1) to request ECN congestion marking behavior in the network that differs from ECT(0) counterbalanced by use of a different IETF-approved congestion response to CE

marks at the sender, e.g., as proposed in [I-D.briscoe-tsvwg-ecn-l4s-id]. This is at variance with RFC 3168's requirement that ECT(0)-marked traffic and ECT(1)-marked traffic not receive different treatment in the network.

Generalized ECN: Use ECN for TCP control packets (i.e., send control packets such as SYN with ECT marking) and for retransmitted packets, e.g., as proposed in [I-D.bagnulo-tsvwg-generalized-ecn]. This is at variance with RFC 3168's prohibition of use of ECN for TCP control packets and retransmitted packets

The scope of this memo is limited to these three areas of experimentation. This memo neither prejudges the outcomes of the proposed experiments nor specifies the experiments in detail. Additional experiments in these areas are possible, e.g., on use of ECN to support deployment of Datacenter TCP (DCTCP) [I-D.ietf-tcpm-dctcp] beyond its current applicablity limitation to data center environments. The purpose of this memo is to remove constraints in standards track RFCs that serve to prohibit these areas of experimentation.

3. ECN Nonce and RFC 3540

As specified in RFC 3168, ECN uses two ECN Capable Transport (ECT) codepoints to indicate that a packet supports ECN, ECT(0) and ECT(1), with the second codepoint used to support ECN nonce functionality to discourage receivers from exploiting ECN to improve their throughput at the expense of other network users, as specified in experimental RFC 3540 [RFC3540].

While the ECN Nonce works as specified, and has been deployed in limited environments, widespread usage in the Internet has not materialized. A study of the ECN behaviour of the Alexa top 1M web servers using 2014 data [Trammell15] found that after ECN was negotiated, none of the 581,711 IPv4 servers tested were using both ECT codepoints, which would have been a possible sign of ECN Nonce usage. Of the 17,028 IPv6 servers tested, 4 set both ECT(0) and ECT(1) on data packets. This might have been evidence of use of the ECN Nonce by these 4 servers, but equally it might have been due to re-marking of the ECN field by an erroneous middlebox or router.

With the emergence of new experimental functionality that depends on use of the ECT(1) codepoint for other purposes, continuing to reserve that codepoint for the ECN Nonce experiment is no longer justified. In addition, other approaches to discouraging receivers from exploiting ECN have emerged, see Appendix B.1 of [I-D.briscoe-tsvwg-ecn-l4s-id]. Therefore, in support of ECN experimentation with the ECT(1) codepoint, this memo:

- Declares that the ECN Nonce experiment [RFC3540] has concluded, and notes the absence of widespread deployment.

- Obsoletes RFC 3540 in order to facilitate experimental use of the ECT(1) codepoint.

- Reclassifies RFC 3540 as Historic to document the ECN Nonce experiment and discourage further implementation of the ECN Nonce.

- Updates RFC 3168 [RFC3168] to remove discussion of the ECN Nonce and use of ECT(1) for that Nonce. The specific text updates are omitted for brevity.

The following guidance on ECT codepoint usage in the ECN field of IP headers from Section 5 of RFC 3168 is relevant when the ECN Nonce is not implemented:

> Protocols and senders that only require a single ECT codepoint SHOULD use ECT(0).

OPEN ISSUE: Change the above requirement in RFC 3168 from SHOULD to MUST towards reserving ECT(1) for experimentation?

4. Updates to RFC 3168

RFC 3168 is a Proposed Standard RFC, so updating RFC 3168 requires publishing a standards track RFC unless a standards process exception is approved by the IESG, e.g., to allow an Experimental RFC to update RFC 3168. In support of the above areas of experimentation, and specifically to avoid multiple uncoordinated requests to the IESG for standards process exceptions, this memo updates RFC 3168 [RFC3168] ito allow changes in the following areas, provided that the changes are documented by an Experimental RFC. It is also possible to change RFC 3168 via publication of another standards track RFC.

4.1. Congestion Response Differences

Section 5 of RFC 3168 specifies that:

> Upon the receipt by an ECN-Capable transport of a single CE packet, the congestion control algorithms followed at the end-systems MUST be essentially the same as the congestion control response to a *single* dropped packet.

In support of Congestion Response Differences experimentation, this memo updates RFC 3168 to allow the congestion control response (including the TCP Sender's congestion control response) to a CE-marked packet to differ from the response to a dropped packet,

provided that the changes from RFC 3168 are documented in an Experimental RFC. The specific change to RFC 3168 is to insert the words "unless otherwise specified by an Experimental RFC" at the end of the sentence quoted above.

RFC 4774 [RFC4774] quotes the above text from RFC 3168 as background, but does not impose requirements based on that text. Therefore no update to RFC 4774 is required to enable this area of experimentation.

4.2. ECT Differences

Section 5 of RFC 3168 specifies that:

> Routers treat the ECT(0) and ECT(1) codepoints as equivalent.

In support of ECT Differences experimentation, this memo updates RFC 3168 to allow routers to treat the ECT(0) and ECT(1) codepoints differently, provided that the changes from RFC 3168 are documented in an Experimental RFC. The specific change to RFC 3168 is to insert the words "unless otherwise specified by an Experimental RFC" at the end of the above sentence.

In support of ECT Differences experimentation, this memo updates RFC 3168 to enable effective endpoint use of ECT(1) for large scale experimentation. The proposed L4S experiment [I-D.briscoe-tsvwg-ecn-l4s-id] significantly increases the CE marking probability for ECT(1)-marked traffic in a fashion that interacts badly with existing sender congestion response functionality because that functionality assumes that a CE-marked packet would have been dropped by the network. If network traffic that uses such a sender congestion response encounters L4S's increased marking probability (and hence rate) at a network bottleneck queue, the resulting traffic throughput is likely to be much less than intended for the level of congestion at the bottleneck queue. To avoid that interaction, this memo reserves ECT(1) for experimentation, initially L4S. The specific update to Section 5 of RFC 3168 is to remove the following text:

> Senders are free to use either the ECT(0) or the ECT(1) codepoint to indicate ECT, on a packet-by-packet basis.
>
> The use of both the two codepoints for ECT, ECT(0) and ECT(1), is motivated primarily by the desire to allow mechanisms for the data sender to verify that network elements are not erasing the CE codepoint, and that data receivers are properly reporting to the sender the receipt of packets with the CE codepoint set, as required by the transport protocol. Guidelines for the senders

> and receivers to differentiate between the ECT(0) and ECT(1) codepoints will be addressed in separate documents, for each transport protocol. In particular, this document does not address mechanisms for TCP end- nodes to differentiate between the ECT(0) and ECT(1) codepoints. Protocols and senders that only require a single ECT codepoint SHOULD use ECT(0).

and replace it with:

> Unless otherwise modified by an Experimental RFC, senders MUST use the ECT(0) codepoint to indicate ECT and MUST NOT use the ECT(1) codepoint to indicate ECT.

ECT Differences experiments SHOULD modify the network behavior for ECT(1)-marked traffic rather than ECT(0)-marked traffic if network behavior for only one ECT codepoint is modified. ECT Differences experiments MUST NOT modify the network behavior for ECT(0)-marked traffic in a fashion that requires changes to sender congestion response to obtain desired network behavior. If an ECT Differences experiment modifies the network behavior for ECT(1)-marked traffic, e.g., CE-marking behavior, in a fashion that requires changes to sender congestion response to obtain desired network behavior, then the Experimental RFC for that experiment MUST specify:

- The sender congestion response to CE marking in the network, and
- Router behavior changes, or the absence thereof, in fowarding CE-marked packets that are part of the experiment.

In addition, until the conclusion of the L4S experiment, use of ECT(1) in IETF RFCs is not appropriate, as the IETF may decide to allocate ECT(1) exclusively for L4S usage if the L4S experiment is successful.

In support of ECT Differences experimentation, this memo also updates RFC 3168 to remove discussion of the ECN Nonce, as noted in Section 3 above.

4.3. Generalized ECN

RFC 3168 prohibits use of ECN for TCP control packets and retransmitted packets in a number of places:

- "To ensure the reliable delivery of the congestion indication of the CE codepoint, an ECT codepoint MUST NOT be set in a packet unless the loss of that packet in the network would be detected by the end nodes and interpreted as an indication of congestion." (Section 5.2)

- "A host MUST NOT set ECT on SYN or SYN-ACK packets." (Section 6.1.1)

- "pure acknowledgement packets (e.g., packets that do not contain any accompanying data) MUST be sent with the not-ECT codepoint." (Section 6.1.4)

- "This document specifies ECN-capable TCP implementations MUST NOT set either ECT codepoint (ECT(0) or ECT(1)) in the IP header for retransmitted data packets, and that the TCP data receiver SHOULD ignore the ECN field on arriving data packets that are outside of the receiver's current window." (Section 6.1.5)

- "the TCP data sender MUST NOT set either an ECT codepoint or the CWR bit on window probe packets." (Section 6.1.6)

In support of Generalized ECN experimentation, this memo updates RFC 3168 to allow the use of ECT codepoints on SYN and SYN-ACK packets, pure acknowledgement packets, window probe packets and retransmissions of packets that were originally sent with an ECT codepoint, provided that the changes from RFC 3168 are documented in an Experimental RFC. The specific change to RFC 3168 is to insert the words "unless otherwise specified by an Experimental RFC" at the end of each sentence quoted above.

In addition, beyond requiring TCP senders not to set ECT on TCP control packets and retransmitted packets, RFC 3168 is silent on whether it is appropriate for a network element, e.g. a firewall, to discard such a packet as invalid. For Generalized ECN experimentation to be useful, middleboxes ought not to do that, therefore RFC 3168 is updated by adding the following text to the end of Section 6.1.1.1 on Middlebox Issues:

> Unless otherwise specified by an Experimental RFC, middleboxes SHOULD NOT discard TCP control packets and retransmitted TCP packets solely because the ECN field in the IP header does not contain Not-ECT.

4.4. Effective Congestion Control is Required

Congestion control remains an important aspect of the Internet architecture [RFC2914]. Any Experimental RFC that takes advantage of this memo's updates to RFC 3168 or RFC 6679 is required to discuss the congestion control implications of the experiment(s) in order to provide assurance that deployment of the experiment(s) does not pose a congestion-based threat to the operation of the Internet.

5. ECN for RTP Updates to RFC 6679

RFC 6679 [RFC6679] specifies use of ECN for RTP traffic; it allows use of both the ECT(0) and ECT(1) codepoints, and provides the following guidance on use of these codepoints in section 7.3.1 :

> The sender SHOULD mark packets as ECT(0) unless the receiver expresses a preference for ECT(1) or for a random ECT value using the "ect" parameter in the "a=ecn-capable-rtp:" attribute.

The ECT Differences area of experimentation increases the potential consequences of using ECT(1) instead of ECT(0), and hence the above guidance is updated by adding the following two sentences:

> Random ECT values MUST NOT be used, as that may expose RTP to differences in network treatment of traffic marked with ECT(1) and ECT(0) and differences in associated endpoint congestion responses, e.g., as proposed in [I-D.briscoe-tsvwg-ecn-l4s-id]. In addition, ECT(0) MUST be used unless otherwise specified in an Experimental RFC.

Section 7.3.3 of RFC 6679 specifies RTP's response to receipt of CE marked packets as being identical to the response to dropped packets:

> The reception of RTP packets with ECN-CE marks in the IP header is a notification that congestion is being experienced. The default reaction on the reception of these ECN-CE-marked packets MUST be to provide the congestion control algorithm with a congestion notification that triggers the algorithm to react as if packet loss had occurred. There should be no difference in congestion response if ECN-CE marks or packet drops are detected.

In support of Congestion Response Differences experimentation, this memo updates this text in a fashion similar to RFC 3168 to allow the RTP congestion control response to a CE-marked packet to differ from the response to a dropped packet, provided that the changes from RFC 6679 are documented in an Experimental RFC. The specific change to RFC 6679 is to insert the words "Unless otherwise specified by an Experimental RFC" and reformat the last two sentences to be subject to that condition, i.e.:

> The reception of RTP packets with ECN-CE marks in the IP header is a notification that congestion is being experienced. Unless otherwise specified by an Experimental RFC:
>
> * The default reaction on the reception of these ECN-CE-marked packets MUST be to provide the congestion control algorithm

with a congestion notification that triggers the algorithm to react as if packet loss had occurred.

* There should be no difference in congestion response if ECN-CE marks or packet drops are detected.

The second sentence of the immediately following paragraph in RFC 6679 requires a related update:

> Other reactions to ECN-CE may be specified in the future, following IETF Review. Detailed designs of such additional reactions MUST be specified in a Standards Track RFC and be reviewed to ensure they are safe for deployment under any restrictions specified.

The update is to change "Standards Track RFC" to "Standards Track RFC or Experimental RFC" for consistency with the first update.

6. ECN for DCCP Updates to RFCs 4341, 4342 and 5622

The specifications of the three DCCP Congestion Control IDs (CCIDs) 2 [RFC4341], 3 [RFC4342] and 4 [RFC5622] contain broadly the same wording as follows:

> each DCCP-Data and DCCP-DataAck packet is sent as ECN Capable with either the ECT(0) or the ECT(1) codepoint set.

This memo updates these sentences in each of the three RFCs as follows:

> each DCCP-Data and DCCP-DataAck packet is sent as ECN Capable. Unless otherwise specified by an Experimental RFC, such DCCP senders SHOULD set the ECT(0) codepoint.

In support of ECT Differences experimentation (as noted in Section 3), this memo also updates all three of these RFCs to remove discussion of the ECN Nonce. The specific text updates are omitted for brevity.

7. Acknowledgements

The content of this draft, including the specific portions of RFC 3168 that are updated draws heavily from [I-D.khademi-tsvwg-ecn-response], whose authors are gratefully acknowledged. The authors of the Internet Drafts describing the experiments have motivated the production of this memo - their interest in innovation is welcome and heartily acknowledged. Colin

Perkins suggested updating RFC 6679 on RTP and provided guidance on where to make the updates.

The draft has been improved as a result of comments from a number of reviewers, including Spencer Dawkins, Gorry Fairhurst, Ingemar Johansson, Naeem Khademi, Mirja Kuehlewind, Karen Nielsen and Michael Welzl. Bob Briscoe's thorough review of an early version of this draft resulted in numerous improvments including addition of the updates to the DCCP RFCs.

8. IANA Considerations

This memo includes no request to IANA.

9. Security Considerations

As a process memo that makes no changes to existing protocols, there are no protocol security considerations.

However, effective congestion control is crucial to the continued operation of the Internet, and hence this memo places the responsibility for not breaking Internet congestion control on the experiments and the experimenters who propose them, as specified in Section 4.4.

Security considerations for the proposed experiments are dicussed in the Internet-Drafts that propose them.

See Appendix B.1 of [I-D.briscoe-tsvwg-ecn-l4s-id] for discussion of alteratives to the ECN Nonce.

10. References

10.1. Normative References

[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>.

[RFC2914] Floyd, S., "Congestion Control Principles", BCP 41, RFC 2914, DOI 10.17487/RFC2914, September 2000, <http://www.rfc-editor.org/info/rfc2914>.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <http://www.rfc-editor.org/info/rfc3168>.

[RFC3540] Spring, N., Wetherall, D., and D. Ely, "Robust Explicit Congestion Notification (ECN) Signaling with Nonces", RFC 3540, DOI 10.17487/RFC3540, June 2003, <http://www.rfc-editor.org/info/rfc3540>.

[RFC4341] Floyd, S. and E. Kohler, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control", RFC 4341, DOI 10.17487/RFC4341, March 2006, <http://www.rfc-editor.org/info/rfc4341>.

[RFC4342] Floyd, S., Kohler, E., and J. Padhye, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)", RFC 4342, DOI 10.17487/RFC4342, March 2006, <http://www.rfc-editor.org/info/rfc4342>.

[RFC5622] Floyd, S. and E. Kohler, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion ID 4: TCP-Friendly Rate Control for Small Packets (TFRC-SP)", RFC 5622, DOI 10.17487/RFC5622, August 2009, <http://www.rfc-editor.org/info/rfc5622>.

[RFC6679] Westerlund, M., Johansson, I., Perkins, C., O'Hanlon, P., and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", RFC 6679, DOI 10.17487/RFC6679, August 2012, <http://www.rfc-editor.org/info/rfc6679>.

10.2. Informative References

[I-D.bagnulo-tsvwg-generalized-ecn] Bagnulo, M. and B. Briscoe, "Adding Explicit Congestion Notification (ECN) to TCP control packets", draft-bagnulo-tsvwg-generalized-ecn-01 (work in progress), July 2016.

[I-D.briscoe-tsvwg-ecn-l4s-id] Schepper, K., Briscoe, B., and I. Tsang, "Identifying Modified Explicit Congestion Notification (ECN) Semantics for Ultra-Low Queuing Delay", draft-briscoe-tsvwg-ecn-l4s-id-02 (work in progress), October 2016.

[I-D.ietf-tcpm-dctcp] Bensley, S., Eggert, L., Thaler, D., Balasubramanian, P., and G. Judd, "Datacenter TCP (DCTCP): TCP Congestion Control for Datacenters", draft-ietf-tcpm-dctcp-04 (work in progress), February 2017.

[I-D.khademi-tcpm-alternativebackoff-ecn]
Khademi, N., Welzl, M., Armitage, G., and G. Fairhurst, "TCP Alternative Backoff with ECN (ABE)", draft-khademi-tcpm-alternativebackoff-ecn-01 (work in progress), October 2016.

[I-D.khademi-tsvwg-ecn-response]
Khademi, N., Welzl, M., Armitage, G., and G. Fairhurst, "Updating the Explicit Congestion Notification (ECN) Specification to Allow IETF Experimentation", draft-khademi-tsvwg-ecn-response-01 (work in progress), July 2016.

[RFC4774] Floyd, S., "Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field", BCP 124, RFC 4774, DOI 10.17487/RFC4774, November 2006, <http://www.rfc-editor.org/info/rfc4774>.

[Trammell15]
Trammell, B., Kuehlewind, M., Boppart, D., Learmonth, I., Fairhurst, G., and R. Scheffenegger, "Enabling Internet-Wide Deployment of Explicit Congestion Notification".

In Proc Passive & Active Measurement (PAM'15) Conference (2015)

Appendix A. Change History

[To be removed before RFC publication.]

Changes from draft-black-tsvwg-ecn-experimentation-00 to -01:

- Section 4.2 - also update RFC 3168 to remove sentence indicating that senders are free to use both ECT codepoints. Add a SHOULD for ECT Differences experiments to use ECT(1).

- Section 5 - only discourage use of random ECT values, but use NOT RECOMMENDED to do so. Consistent use of ECT(1) without using ECT(0) is ok. Mention possible changes in endpoint response.

- Add more Acknowledgements and Change History

- Additional editorial changes.

Changes from draft-black-tsvwg-ecn-experimentation-01 to -02:

- Add DCCP RFC updates and one missing RFC 3168 update (probe packets).

- Discourage RTP usage of ECT(1).
- Strengthen text on lack of ECN Nonce deployment.
- Cross-reference the L4S draft appendix that discusses ECN Nonce alternatives.
- Additional editorial changes.

Changes from draft-black-tsvwg-ecn-experimentation-02 to -03:

- Clarify that "SHOULD use ECT(0)" guidance from RFC 3168 is about IP headers.
- Add a "SHOULD NOT" requirement that middleboxes not discard TCP control packets, etc. solely because they use ECN.
- Switch to pre-5378 boilerplate, due to vintage of RFCs being updated.
- Additional editorial changes.

Changes from draft-black-tsvwg-ecn-experimentation-03 to -04:

- Use "Congestion Response Differences" as name of experimentation area instead of "Alternative Backoff" to avoid confusion with specific experiment.
- Change ECT(1) requirement to "MUST NOT use unless otherwise specified by an Experimental RFC" This resulted in extensive changes to Section 4.2.
- Clean up and tighten language requiring all congestion responses to be IETF-approved
- Additional editorial changes.

Initial WG draft, draft-ietf-tsvwg-ecn-experimentation-00, has same contents as draft-black-tsvwg-ecn-experimentation-04.

Changes from draft-ietf-tsvwg-ecn-experimentation-00 to -01:

- Add mention of DCTCP as another protocol that could benefit from ECN experimentation (near end of Section 2).

Author's Address

David Black
Dell EMC
176 South Street
Hopkinton, MA 01748
USA

Email: david.black@dell.com