Network Working Group
Internet-Draft
Intended status: Informational
Expires: May 3, 2016

G. Fairhurst
University of Aberdeen
M. Welzl
University of Oslo

October 31, 2015

The Benefits of using Explicit Congestion Notification (ECN)
draft-ietf-aqm-ecn-benefits-07

Abstract

The goal of this document is to describe the potential benefits when applications use a transport that enables Explicit Congestion Notification (ECN). The document outlines the principal gains in terms of increased throughput, reduced delay and other benefits when ECN is used over a network path that includes equipment that supports Congetsion Experienced (CE) marking. It also discusses challenges for successful deployment of ECN. It does not propose new algorithms to use ECN, nor does it describe the details of implementation of ECN in endpoint devices (Internet hosts), routers or 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 May 3, 2016.

Table of Contents

1. Introduction

Internet Transports (such as TCP and SCTP) are implemented in endpoints (Internet hosts) and are designed to detect and react to network congestion. Congestion may be detected by loss of an IP packet or, if Explicit Congestion Notification (ECN) [RFC3168] is enabled, by the reception of a packet with a Congestion Experienced (CE) marking in the IP header. Both of these are treated by transports as indications of congestion. ECN may also be enabled by

other transports: UDP applications that provide congestion control may enable ECN when they are able to correctly process the ECN signals [ID.RFC5405.bis] (e.g., ECN with RTP [RFC6679]).

Active Queue Management (AQM) [RFC7567] is a class of techniques that can be used by network devices (a router, middlebox, or other device that forwards packets through the network) to manage the size of queues in network buffers.

A network device that does not support AQM typically uses a drop-tail policy to drop excess IP packets when its queue becomes full. The discard of packets is a signal to the end-to-end treated by transports as an indications of congestion on the network path being used. (Although packet loss may also occur for various other reasons, e.g., link corruption, receiver-overrun, endpoints can not differentiate types of loss and therefore need to regard all loss as potentially caused by congestion.) Observed loss therefore results in a congestion control reaction by the transport to reduce the maximum rate permitted by the sending endpoint.

The packet header of all IPv4 and IPv6 packets IP packet carry an ECN field [RFC3168]. This field may be set to one of four values shown in Table 1. The not-ECT codepoint '00' indicates a packet that is not using ECN. The ECT(0) codepoint '01' and the ECT(1) codepoint '10' both indicate that the transport protocol using the IP layer supports the use of ECN. The CE codepoint '11' is set by an ECN-capable network device to indicate congestion to the transport endpoint.

ECN FIELD		Name
0	0	Not-ECT
0	1	ECT(1)
1	0	ECT(0)
1	1	CE

Table 1: The ECN Field in the IP Packet Header (based on [RFC3168]).

When an application uses a transport that enables use of ECN [RFC3168], the transport layer sets the ECT(0) or ECT(1) codepoint in the IP header of packets that it sends. This indicates to network devices that they may mark, rather than drop the ECN-capable IP packets. An ECN-capable network device can then signal incipient congestion (network queueing) at a point before a transport experiences congestion loss or high queuing delay. The marking is

generally performed as the result of various AQM algorithms [RFC7567], where the exact combination of AQM/ECN algorithms does not need to be known by the transport endpoints.

Since ECN makes it possible for the network to signal the presence of incipient congestion without incurring packet loss, it lets the network deliver some packets to an application that would otherwise have been dropped if the application or transport did not support ECN. 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 receiving CE-marking instead of dropping packets. Several benefits reduce latency (e.g., reduced Head-of-Line Blocking).

The focus of the document is on usage of ECN by transport and application layer flows, not its implementation in endpoint hosts, or in routers and other network devices.

1.1. Terminology

The following terms are used:

AQM: Active Queue Management.

CE: Congestion Experienced, a codepoint value '11' marked in the ECN field of the IP packet header.

ECN-capable receiving endpoint: An endpoint that passes the ECN field to the relevant transport protocol that can process the values and generate appropriate feedback to control the rate of the sending endpoint.

ECN-capable IP Packet : A packet where the ECN field is set to a non-zero ECN value (i.e., with a ECT(0), ECT(1), or the CE codepoint).

ECN-capable network device : An ECN-capable network device may forward, drop, or queue an ECN-capable packet and may choose to CE-mark this packet when there is incipient congestion.

ECN field: A 2-bit field specified for use explicit congestion signalling in the IPv4 and IPv6 packet headers.

Endpoint: An Internet host that terminates a transport protocol connection across an Internet path.

Incipient Congestion: The detection of congestion when it is starting, perhaps by a network device noting that the arrival rate exceeds the forwarding rate.

Network device: A router, middlebox, or other device that forwards IP packets through the network.

non-ECN-capable: A network device or endpoint that does not interpret the ECN field. Such a device is not permitted to change the ECN codepoint.

not-ECN-capable IP Packet: An IP packet with the ECN field set to a value of zero ('00'). A not-ECN-capable packet may be forwarded, dropped or queued by a network device.

2. Benefit of using ECN to avoid Congestion Loss

An ECN-capable network device is expected to CE-mark an ECN-capable IP packet when an AQM method detects incipient congestion, rather than to drop the packet [RFC7567]. An application can benefit from this marking in several ways:

2.1. Improved Throughput

ECN seeks to avoid the inefficiency of dropping data that has already made it across at least part of the network path.

ECN can improve the throughput of an application, although this increase in throughput 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 would be 0.02 (i.e., 1 in 50 packets). This translates into an approximate 2% throughput gain if ECN is enabled. (Note that in heavy congestion, packet loss may be unavoidable with, or without, ECN.)

2.2. Reduced Head-of-Line Blocking

Many Internet transports provide in-order delivery of received data segments to the applications they support. For these applications, use of ECN can reduce the delay that can result when these applications experience packet loss.

Packet loss may occur for various reasons. One cause arises when an AQM scheme drops a packet as a signal of incipient congestion.

Whatever the cause of loss, a missing packet needs to trigger a congestion control response. A reliable transport also triggers retransmission to recover the lost data. For a transport providing in-order delivery, this requires that the transport receiver stalls (or waits) for all data that was sent ahead of a lost segment to be correctly received before it can forward any later data to the application. A loss therefore creates a delay of at least one RTT after a loss event before data can be delivered to an application. We call this Head-of-Line (HOL) blocking. This is the usual requirement for TCP and SCTP. (PR-SCTP [RFC3758], UDP [RFC0768][ID.RFC5405.bis], and DCCP [RFC4340] provide a transport that does not provide re-ordering).

By enabling ECN, a transport continues to receive in-order data when there is incipient congestion, and can pass this data to the receiving application. Use of ECN avoids the additional reordering delay in a reliable transport. The sender still needs to make an appropriate congestion-response to reduce the maximum transmission rate for future traffic, which usually will require a reduction in the sending rate [ID.RFC5405.bis].)

2.3. Reduced Probability of RTO Expiry

Some patterns of packet loss can result in a Retransmission Time Out (RTO), which causes a sudden and significant change in the allowed rate at which a transport/application can forward packets. Because ECN provides an alternative to drop for network devices to signal incipient congestion, this can reduce the probability of loss and hence reduce the likelihood of RTO expiry.

Internet transports/applications generally use a RTO timer as a last resort to detect and recover loss [ID.RFC5405.bis] [RFC5681]). Specifically, a RTO timer detects loss of a packet that is not followed by other packets, such as at the end of a burst of data segments or when an application 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). This loss of the last segment (or last few segments) of a traffic burst is also known as a "tail loss". Standard transport recovery methods, such as Fast Recovery ([RFC5681], are often unable to recover from a tail loss. This is because the endpoint receiver is unaware that the lost segments were actually sent, and therefore generates no feedback [Fla13]. Retransmission of these segments therefore relies on expiry of a transport retransmission timer. This timer is also used to detect a lack of forwarding along a path. Expiry of the RTO therefore results in the consequent loss of state about the network path being used. This typically includes resetting path estimates such as the RTT, re-

initialising the congestion window, and possibly updates to other transport state. This can reduce the performance of the transport until it again adapts to the path.

An ECN-capable network device cannot eliminate the possibility of tail loss, because a drop may occur due to a traffic burst exceeding the instantaneous available capacity of a network buffer or as a result of the AQM algorithm (overload protection mechanisms, etc [RFC7567]). However, an ECN-capable network device that observes incipient congestion may be expected to buffer the IP packets of an ECN-capable flow and set a CE-mark in one or more packet(s), rather than triggering packet drop. Setting a CE-mark signals incipient congestion without forcing the transport/application to enter retransmission timeout. This reduces application-level latency and can improve the throughput for applications that send intermittent bursts of data.

The benefit of avoiding retransmission loss is expected to be significant when ECN is used on TCP SYN/ACK packets [RFC5562] where the RTO interval may be large because TCP cannot base the timeout period on prior RTT measurements from the same connection.

2.4. Applications that do not Retransmit Lost Packets

A transport that enables ECN can receive timely congestion signals without the need to retransmit packets each time it receives a congestion signal.

Some latency-critical applications do not retransmit lost packets, yet may be able to adjust their sending rate following detection of incipient congestion. Examples of such applications include UDP-based services that carry Voice over IP (VoIP), interactive video, or real-time data. The performance of many such applications degrades rapidly with increasing packet loss and the transport/application may therefore employ mechanisms (e.g., packet forward error correction, data duplication, or media codec error concealment) to mitigate the immediate effect of congestion loss on the application. Some mechanisms consume additional network capacity, some require additional processing and some contribute additional path latency when congestion is experienced. By decoupling congestion control from loss, ECN can allow transports that support these applications to reduce their rate before the application experiences loss from congestion. This can reduce the negative impact of triggering loss-hiding mechanisms with a direct positive impact on the quality experienced by the users of these applications.

2.5. Making Incipient 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 allowing information to be collected about the presence of incipient congestion.

Recording the presence of CE-marked packets can provide information about the current congestion level experienced on a network path. A network flow that only experiences CE-marking and no loss implies that the sending endpoint is experiencing only congestion. A network flow may also experience loss (e.g., due to queue overflow, AQM methods that protect other flows, link corruption or loss in middleboxes). When a mixture of CE-marking and packet loss is experienced, transports and measurements need to assume there is congestion [RFC7567]. An absence of CE-marks therefore does not indicate a path has not experienced congestion.

The reception of CE-marked packets can be used to monitor the level of congestion by a transport/application or a network operator. For example, ECN measurements are used by Congestion Exposure (ConEx) [RFC6789]. In contrast, metering packet loss is harder.

2.6. Opportunities for new Transport Mechanisms

ECN can enable design and deployment of new algorithms in network devices and Internet transports. Internet transports need to regard both loss and CE-marking as an indication of congestion. However, while the amount of feedback provided by drop ought naturally to be minimized, this is not the case for ECN. In contrast, an ECN-Capable network device could provide richer (more frequent and fine-grained) indication of its congestion state to the transport.

All ECN-capable receiving endpoints need to provide feedback to the transport sender to indicate that CE-marks have been received.[RFC3168] provides one method that signals once each round trip time that CE-marked packets have been received.

A receiving endpoint may provide more detailed feedback to the congestion controller at the sender (e.g., describing the set of received ECN codepoints, or indicating each received CE-marked packet). Precise feedback about the number of CE-marks encountered is supported by the Real Time Protocol (RTP) when used over UDP [RFC6679] and has been proposed for SCTP [ST14] and TCP [ID.Acc.ECN].

More detailed feedback is expected to enable evolution of transport protocols allowing the congestion control mechanism to make a more appropriate decision on how to react to congestion. Designers of

transport protocols need to consider not only how network devices CE-mark packets, but also how the control loop in the application/ transport reacts to reception of these CE-marked packets.

Benefit has been noted when packets are CE-marked early using an instantaneous queue, and if the receiving endpoint provides feedback about the number of packet marks encountered, an improved sender behavior has been shown to be possible, e.g, Datacenter TCP (DCTCP) [AL10]. DCTCP is targeted at controlled environments such as a datacenter. This is work-in-progress and it is currently unknown whether or how such behaviour could be safely introduced into the Internet. Any update to an Internet transport protocol requires careful consideration of the robustness of the behaviour when working with endpoints or network devices that were not designed for the new congestion reaction.

3. Network Support for ECN

For an application to use ECN requires that the endpoints first enable ECN within the transport being used, but also for all network devices along the path to at least forward IP packets that set a non-zero ECN codepoint.

ECN can be deployed both in the general Internet and in controlled environments:

- ECN can be incrementally deployed in the general Internet. The IETF has provided guidance on configuration and usage in [RFC7567].

- ECN may be deployed within a controlled environment, for example within a data centre or within a well-managed private network. This use of ECN may be tuned to the specific use-case. An example is DCTCP [AL10] [ID.DCTCP].

Early experience of using ECN across the general Internet encountered a number of operational difficulties when the network path either failed to transfer ECN-capable packets or inappropriately changed the ECN codepoints [BA11]. A recent survey reported a growing support for network paths to pass ECN codepoints [TR15].

The remainder of this section identifies what is needed for network devices to effectively support ECN.

3.1. The ECN Field

The current IPv4 and IPv6 specifications assign usage of 2 bits in the IP header to carry the ECN codepoint. This 2-bit field was reserved in [RFC2474] and assigned in [RFC3168].

[RFC4774] discusses some of the issues in defining alternate semantics for the ECN field, and specifies requirements for a safe coexistence in an Internet that could include routers that do not understand the defined alternate semantics.

Some network devices were configured to use a routing hash that included the set of 8 bits forming the now deprecated Type of Service (ToS) field [RFC1349]. The present use of this field assigns 2 of these bits to carry the ECN field. This is incompatible with use in a routing hash, because it could lead to IP packets that carry a CE-mark being routed over a different path to those packets that carried an ECT mark. The resultant reordering would impact the performance of transport protocols (such as TCP or SCTP) and UDP-based applications that are senstive to reordering. A network device that conforms to this older specification needs to be updated to the current specifications [RFC2474] to support ECN. Configuraton of network devices must note that the ECN field may be updated by any ECN-capable network device along a path.

3.2. Forwarding ECN-Capable IP Packets

Not all network devices along a path need to be ECN-capable (i.e., perform CE-marking). However, all network devices need to be configured not to drop packets solely because the ECT(0) or ECT(1) codepoints are used.

Any network device that does not perform CE-marking of an ECN-capable packet can be expected to drop these packets under congestion. Applications that experience congestion at these network devices do not see any benefit from enabling ECN. However, they may see benefit if the congestion were to occur within a network device that did support ECN.

3.3. Enabling ECN in Network Devices

Network devices should use an AQM algorithm that CE-marks ECN-capable traffic when making decisions about the response to congestion [RFC7567]. An ECN method should set a CE-mark on ECN-capable packets in the presence of incipient congestion. A CE-marked packet will be interpreted as an indication of incipient congestion by the transport endpoints.

There is opportunity to design an AQM method for an ECN-capable network device that differs from an AQM method designed to drop packets. [RFC7567] states that the network device should allow this behaviour to be configurable.

[RFC3168] describes a method in which a network device sets the CE-mark at the time that the network device would otherwise have dropped the packet. While it has often been assumed that network devices should CE-mark packets at the same level of congestion at which they would otherwise have dropped them, [RFC7567] recommends that network devices allow independent configuration of the settings for AQM dropping and ECN marking. Such separate configuration of the drop and mark policies is supported in some network devices.

3.4. Co-existance of ECN and non-ECN flows

Network devices need to be able to forward all IP flows and provide appropriate treatment for both ECN and non-ECN traffic.

The design considerations for an AQM scheme supporting ECN needs to consider the impact of queueing during incipient congestion. For example, a simple AQM scheme could choose to queue ECN-capable and non-ECN capable flows in the same queue with an ECN scheme that CE-mark packets during incipient congestion. The CE-marked packets that remain in the queue during congestion can continue to contribute to queueing delay. In contrast, non-ECN-capable packets would normally be dropped by an AQM scheme under incipient congestion. This difference in queueing is one motivation for consideration of more advanced AQM schemes, and may provide an incentive for enabling flow isolation using scheduling [RFC7567]. The IETF is defining methods to evaluate the suitability of AQM schemes for deployment in the general Internet [ID.AQM.eval].

3.5. Bleaching and Middlebox Requirements to deploy ECN

Network devices should not be configured to change the ECN codepoint in the packets that they forward, except to set the CE-codepoint to signal incipient congestion.

Cases have been noted where an endpoint sends a packet with a non-zero ECN mark, but the packet is received by the remote endpoint with a zero ECN codepoint [TR15]. This could be a result of a policy that erases or "bleaches" the ECN codepoint values at a network edge (resetting the codepoint to zero). Bleaching may occur for various reasons (including normalising packets to hide which equipment supports ECN). This policy prevents use of ECN by applications.

When ECN-capable IP packets, marked as ECT(0) or ECT(1), are remarked to non-ECN-capable (i.e., the ECN field is set to zero codepoint), this could result in the packets being dropped by ECN-capable network devices further along the path. This eliminates the advantage of using of ECN.

A network device must not change a packet with a CE mark to a zero codepoint, if the network device decides not to forward the packet with the CE-mark, it has to instead drop the packet and not bleach the marking. This is because a CE-marked packet has already received ECN treatment in the network, and remarking it would then hide the congestion signal from the receiving endpoint. This eliminates the benefits of ECN. It can also slow down the response to congestion compared to using AQM, because the transport will only react if it later discovers congestion by some other mechanism.

Prior to RFC2474, a previous usage assigned the bits now forming the ECN field as a part of the now deprecated Type of Service (ToS) field [RFC1349]. A network device that conforms to this older specification was allowed to remark or erase the ECN codepoints, and such equipment needs to be updated to the current specifications to support ECN.

3.6. Tunneling ECN and the use of ECN by Lower Layer Networks

Some networks may use ECN internally or tunnel ECN (e.g., for traffic engineering or security). These methods need to ensure that the ECN-field of the tunnel packets is handled correctly at the ingress and egress of the tunnel. Guidance on the correct use of ECN is provided in [RFC6040].

Further guidance on the encapsulation and use of ECN by non-IP network devices is provided in [ID.ECN-Encap].

4. Using ECN across the Internet

A receiving endpoint needs to report the loss it experiences when it uses loss-based congestion control. So also, when ECN is enabled, a receiving endpoint must correctly report the presence of CE-marks by providing a mechanism to feed this congestion information back to the sending endpoint , [RFC3168], [ID.RFC5405.bis], enabling the sender to react to experienced congestion. This mechanism needs to be designed to operate robustly across a wide range of Internet path characteristics. This section describes partial deployment, how ECN-enabled endpoints can continue to work effectively over a path that experiences misbehaving network devices or when an endpoint does not correctly provide feedback of ECN congestion information.

4.1. Partial Deployment

Use of ECN is negotiated between the endpoints prior to using the mechanism.

ECN has been designed to allow incremental partial deployment [RFC3168]. Any network device can choose to use either ECN or some other loss-based policy to manage its traffic. Similarly, transport/application negotiation allows senders and receiving endpoints to choose whether ECN will be used to manage congestion for a particular network flow.

4.2. Detecting whether a Path Really Supports ECN

Internet transport and applications need to be robust to the variety and sometimes varying path characteristics that are encountered in the general Internet. They need to monitor correct forwarding of ECN over the entire path and duration of a session.

To be robust, applications and transports need to be designed with the expectation of heterogeneous forwarding (e.g., where some IP packets are CE-marked by one network device, and some by another, possibly using a different AQM algorithm, or when a combination of CE-marking and loss-based congestion indications are used. ([ID.AQM.eval] describes methodologies for evaluating AQM schemes.)

A transport/application also needs to be robust to path changes. A change in the set of network devices along a path could impact the ability to effectively signal or use ECN across the path, e.g., when a path changes to use a middlebox that bleaches ECN codepoints (see Section 3.5).

A sending endpoint can check that any CE-marks applied to packets received over the path are indeed delivered to the remote receiving endpoint and that appropriate feedback is provided. (This could be done by a sender setting known a CE codepoint for specific packets in a network flow and then checking whether the remote endpoint correctly reports these marks [ID.Fallback], [TR15].) If a sender detects persistent misuse of ECN, it needs to fall back to using loss-based recovery and congestion control. Guidance on a suitable tranport reaction is provided in [ID.Fallback].

4.3. Detecting ECN Receiver Feedback Cheating

Appropriate feedback requires that the endpoint receiver does not try to conceal reception of CE-marked packets in the ECN feedback information provided to the sending endpoint [RFC7567]. Designers of applications/transports are therefore encouraged to include

mechanisms that can detect this misbehavior. If a sending endpoint detects that a receiver is not correctly providing this feedback, it needs to fall back to using loss-based recovery instead of ECN.

5. Summary: Enabling ECN in Network Devices and Hosts

This section summarises the benefits of deploying and using ECN within the Internet. It also provides a list of prerequisites to achieve ECN deployment.

Application developers should where possible use transports that enable ECN. Applications that directly use UDP need to provide support to implement the functions required for ECN [ID.RFC5405.bis]. Once enabled, an application that uses a transport that supports ECN will experience the benefits of ECN as network deployment starts to enable ECN. The application does not need to be rewritten to gain these benefits. Table 2 summarises the key benefits.

Section	Benefit
2.1	Improved throughput
2.2	Reduced Head-of-Line blocking
2.3	Reduced probability of RTO Expiry
2.4	Applications that do not retransmit lost packets
2.5	Making incipient congestion visible
2.6	Opportunities for new transport mechanisms

Table 2: Summary of Key Benefits

Network operators and people configuring network devices should enable ECN [RFC7567].

Prerequisites for network devices (including IP routers) to enable use of ECN include:

- A network device that updates the ECN field in IP packets must use IETF-specified methods (see Section 3.1).
- A network device may support alternate ECN semantics (see Section 3.1).
- A network device must not choose a different network path solely because a packet carries has a CE-codepoint set in the ECN Field, CE-marked packets need to follow the same path as packets with an ECT(0) or ECT(1) codepoint (see Section 3.1).Network devices need

to be configured not to drop packets solely because the ECT(0) or ECT(1) codepoints are used (see Section 3.2).

- A network device must not change a packet with a CE mark to a not-ECN-capable codepoint ('00'), if the network device decides not to forward the packet with the CE-mark, it has to instead drop the packet and not bleach the marking (see Section 3.5).

- An ECN-capable network device should correctly update the ECN codepoint of ECN-capable packets in the presence of incipient congestion (see Section 3.3).

- Network devices need to be able to forward both ECN-capable and not-ECN-capable flows (see Section 3.4).

Prerequisites for network endpoints to enable use of ECN include:

- An application should use an Internet transport that can set and receive ECN marks (see Section 4).

- An ECN-capable transport/application must return feedback indicating congestion to the sending endpoint and perform an appropriate congestion response (see Section 4).

- An ECN-capable transport/application should detect paths where there is there is persistent misuse of ECN and fall back to not sending ECT(0) or ECT(1) (see Section 4.2).

- Designers of applications/transports are encouraged to include mechanisms that can detect and react appropriately to misbehaving receivers that fail to report CE-marked packets (see Section 4.3).

6. 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, Colin Perkins, Richard Scheffenegger, Dave Taht, Wes Eddy, Fred Baker, Mikael Abrahamsson, Mirja Kuehlewind, John Leslie, and other members of the TSVWG and AQM working groups.

7. IANA Considerations

XX RFC Ed - PLEASE REMOVE THIS SECTION XXX

This memo includes no request to IANA.

8. Security Considerations

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

9. Revision Information

XXX RFC-Ed please remove this section prior to publication.

Revision 00 was the first WG draft.

Revision 01 includes updates to complete all the sections and a rewrite to improve readability. Added section 2. Author list reversed, since Gorry has become the lead author. Corrections following feedback from Wes Eddy upon review of an interim version of this draft.

Note: Wes Eddy raised a question about whether discussion of the ECN Pitfalls could be improved or restructured - this is expected to be addressed in the next revision.

Revision 02 updates the title, and also the description of mechanisms that help with partial ECN support.

We think this draft is ready for wider review. Comments are welcome to the authors or via the IETF AQM or TSVWG mailing lists.

Revision 03 includes updates from the mailing list and WG discussions at the Dallas IETF meeting.

The section "Avoiding Capacity Overshoot" was removed, since this refers primarily to an AQM benefit, and the additional benefits of ECN are already stated. Separated normative and informative references

Revision 04 (WG Review during WGLC)

Updated the abstract.

Added a table of contents.

Addressed various (some conflicting) comments during WGLC with new text.

The section on Network Support for ECN was moved, and some suggestions for rewording sections were implemented.

Decided not to remove section headers for 2.1 and 2.2 - to ensure the document clearly calls-out the benefits.

Updated references. Updated text to improve consistency of terms and added definitions for key terms.

Note: The group suggested this document should not define recommendations for end hosts or routers, but simply state the things needs to enable deployment to be successful.

Revision 05 (after WGLC comments)

Updated abstract to avoid suggesting that this describes new methods for deployment.

Added ECN-field definition, and sorted terms in order.

Added an opening para to each "benefit" to say what this is. Sought to remove redundancy between sections.

Added new section on Codepoints to avoid saying the same thing twice.

Reworked sections 3 and 4 to clarify discussion and to remove unnecessary text.

Reformatted Summary to refer to sections describing things, rather than appear as a list of new recommendations. Reordered to match the new document order.

Note: This version expects an update to RFC5405.bis that will indicate UDP ECN requirements (normative).

Revision 06

Corrections from Miria.

Revision 07

Update to include IESG feedback from: Spencer, Dan, Benoit, Joel. Corrected Non-ECN to Not-ECN where appropriate, added table of codepoints, clarified sentences describing "conservative" behaviour,

added requirement to not do ToS-based routing (Junos enhanced hash), etc. Ammended Acknowledgments section.

10. References

10.1. Normative References

[ID.RFC5405.bis] Eggert, Lars., Fairhurst, Gorry., and Greg. Shepherd, "Unicast UDP Usage Guidelines", 2015.

[RFC2474] "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers".

[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>.

[RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, DOI 10.17487/RFC6040, November 2010, <http://www.rfc-editor.org/info/rfc6040>.

[RFC7567] Baker, F. and G. Fairhurst, "IETF Recommendations Regarding Active Queue Management", Internet-draft draft-ietf-aqm-recommendation-06, October 2014.

10.2. 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.

[BA11] Bauer, Steven., Beverly, Robert., and Arthur. Berger, "Measuring the State of ECN Readiness in Servers, Clients, and Routers, ACM IMC", 2011.

[Fla13] Flach, Tobias., Dukkipati, Nandita., Terzis, Andreas., Raghavan, Barath., Cardwell, Neal., Cheng, Yuchung., Jain, Ankur., Hao, Shuai., Katz-Bassett, Ethan., and Ramesh. Govindan, "Reducing web latency: the virtue of gentle aggression.", SIGCOMM 2013, October 2013.

[ID.Acc.ECN] Briscoe, Bob., Scheffeneger, Richard., and Mirja. Kuehlewind, "More Accurate ECN Feedback in TCP, Work-in-Progress".

[ID.AQM.eval]
Kuhn, Nicolas., Natarajan, Preethi., Ros, David., and Naeem. Khademi, "AQM Characterization Guidelines (Work-in-progress, draft-ietf-aqm-eval-guidelines)", 2015.

[ID.DCTCP]
Bensley, S., Eggert, Lars., and D. Thaler, "Microsoft's Datacenter TCP (DCTCP): TCP Congestion Control for Datacenters (Work-in-progress, draft-bensley-tcpm-dctcp)", 2015.

[ID.ECN-Encap]
Briscoe, B., Kaippallimalil, J., and P. Thaler, "Guidelines for Adding Congestion Notification to Protocols that Encapsulate IP", Internet-draft, IETF work-in-progress draft-ietf-tsvwg-ecn-encap-guidelines.

[ID.Fallback]
Kuehlewind, Mirja. and Brian. Trammell, "A Mechanism for ECN Path Probing and Fallback, draft-kuehlewind-tcpm-ecn-fallback, Work-in-Progress".

[RFC0768] Postel, J., "User Datagram Protocol", 1980.

[RFC1349] "Type of Service in the Internet Protocol Suite".

[RFC3649] Floyd, S., "HighSpeed TCP for Large Congestion Windows", RFC 3649, DOI 10.17487/RFC3649, December 2003, <http://www.rfc-editor.org/info/rfc3649>.

[RFC3758] Stewart, R., Ramalho, M., Xie, Q., Tuexen, M., and P. Conrad, "Stream Control Transmission Protocol (SCTP) Partial Reliability Extension", RFC 3758, DOI 10.17487/RFC3758, May 2004, <http://www.rfc-editor.org/info/rfc3758>.

[RFC4340] Kohler, E., Handley, M., and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", RFC 4340, DOI 10.17487/RFC4340, March 2006, <http://www.rfc-editor.org/info/rfc4340>.

[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>.

[RFC5562] Kuzmanovic, A., Mondal, A., Floyd, S., and K. Ramakrishnan, "Adding Explicit Congestion Notification (ECN) Capability to TCP's SYN/ACK Packets", RFC 5562, DOI 10.17487/RFC5562, June 2009, <http://www.rfc-editor.org/info/rfc5562>.

[RFC5681] Allman, M., Paxson, V., and E. Blanton, "TCP Congestion Control", RFC 5681, DOI 10.17487/RFC5681, September 2009, <http://www.rfc-editor.org/info/rfc5681>.

[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>.

[RFC6789] Briscoe, B., Ed., Woundy, R., Ed., and A. Cooper, Ed., "Congestion Exposure (ConEx) Concepts and Use Cases", RFC 6789, DOI 10.17487/RFC6789, December 2012, <http://www.rfc-editor.org/info/rfc6789>.

[ST14] Stewart, R., Tuexen, M., and X. Dong, "ECN for Stream Control Transmission Protocol (SCTP)", Internet-draft draft-stewart-tsvwg-sctpecn-05.txt, January 2014.

[TR15] Tranmmel, Brian., Kuehlewind, Mirja., Boppart, Damiano, Learmonth, Iain., and Gorry. Fairhurst, "Enabling internet-wide deployment of Explicit Congestion Notification Tramwell, B., Kuehlewind, M., Boppart, D., Learmonth, I., Fairhurst, G. & Scheffnegger, Passive and Active Measurement Conference (PAM)", March 2015.

Authors' Addresses

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

Email: gorry@erg.abdn.ac.uk

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

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