Internet Engineering Task Force K. K. Ramakrishnan, AT&T Research
INTERNET DRAFT Sally Floyd, ACIRI
draft-ietf-mpls-ecn-00.txt Bruce Davie, Cisco
June 1999
Expires: Dec 1999

A Proposal to Incorporate ECN in MPLS

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

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

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

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

Abstract

We propose the addition of ECN to MPLS switching. ECN enables end-end congestion control without necessarily depending on packet loss as the sole indicator of congestion. The proposal suggests having a single bit in the MPLS header to indicate ECN, and simple signaling at the time of setting up the LSP (Label Switched Path) for indication of ECN capability. This allows for incorporation of ECN in MPLS in a coordinated manner with IP and also in a backwards compatible fashion.

1. Introduction.

ECN (Explicit Congestion Notification) [ECN] has been proposed as an addition to IP to provide an indication of congestion. With the addition of active queue management (e.g., RED [RFC2309]) to the Internet infrastructure, where routers detect congestion before the queue overflows, routers are no longer limited to packet drops as an indication of congestion. Routers could instead set a Congestion Experienced (CE) bit in the IP header of packets from ECN-capable transport protocols.

Active queue management mechanisms in the router may use one of several methods for indicating congestion to end-nodes. One is to use packet drops, as is currently done. However, active queue management allows the router to separate policies of queueing or dropping packets from the policies for indicating congestion. Thus, active queue management allows routers to use the Congestion Experienced (CE) bit in a packet header as an indication of congestion, instead of relying solely on packet drops.

Active queue management not only avoids packet losses for congestion indication, but also attempts to control the average queue sizes at routers by using ECN or packet drops to provide an indication of the onset of congestion. With the cooperation of transport protocols, the indication of incipient congestion may be used to minimize significant queue build-up and thus reduce queueing delays, variability in queueing delay and additional packet losses. With ECN-capable routers and transport protocols, packet drops are not required for these indications of congestion.

Applications that are sensitive to the delay or loss of one or more individual packets are expected to benefit from the use of ECN. Interactive traffic such as telnet, web-browsing, and transfer of audio and video data can be sensitive to packet losses (using an unreliable data delivery transport such as UDP) or to the increased latency of the packet caused by the need to retransmit the packet after a loss (for reliable data delivery such as TCP). The use of ECN by end-systems and participation of intermediate nodes in the network in ECN would potentially help these applications.

1.1. Motivation for use of ECN with MPLS

Many of the features of MPLS, such as traffic engineering (to take advantage of multiple paths and balance the load on them), explicit routing and QoS routing are motivated by the desire to improve the overall performance of an IP network. MPLS also aims to support the QoS models that are available for IP, such as Integrated Services and Differentiated Services.

Given the motivation to provide better performance and QoS assurances, we believe it is desirable that we utilize techniques that help manage queueing better, and minimize losses. Label Switched Routers (LSRs) are already expected to incorporate active queue management methods such as RED. As a result, the incremental effort involved in the addition of ECN is likely to be small in many cases. We believe the benefits from ECN of better overall performance for a wide range of applications because of reduced packet loss (especially those using non-TCP transports) and reduced queueing delay is sufficiently significant. Furthermore, there seems to be no reasons for LSRs to lack this capability as it becomes available in other (non-label switched) IP routers.

2. Overview of ECN

This section briefly outlines the addition of ECN to the IP protocol as specified in RFC 2481. RFC 2481 specifies two bits in the ECN field in the IP header, the ECN-Capable Transport (ECT) bit and the Congestion Experienced (CE) bit. The ECT bit is set by the data sender to indicate that the end-points of the transport protocol are ECN-capable. The CE bit is set by the router to indicate congestion to the end nodes. Routers that have a packet arriving at a full queue would drop the packet, just as they do now.

RFC 2481 also specifies the use of ECN in the TCP transport protocol. For an ECN-capable TCP connection, when the TCP receiver receives a data packet with the CE bit set, the receiver conveys that information to the TCP sender using a bit in the TCP header. TCP senders react to the congestion indication by decreasing their congestion window. Thus, the early indication of congestion allows the sender TCP to reduce the load placed on the network, without the need to drop a packet. Because the use of ECN at the transport level is not affected by the MPLS header, this aspect of ECN is not discussed further in this document.

2.1. Opportunities to take Advantage of ECN

In the past, it has often been the desire of datalink layers to provide a congestion indication to the end-systems, so that end-system transport protocols may react by reducing the load. However, even though Frame-Relay and ATM have had an indicator for congestion indication, the match with the ECN indication has not been ideal. In these cases, marking the congestion indication was left to implementation or was based on instantaneous queue occupancy. This was the case with both Forward Explicit Congestion Notification (FECN) in Frame-Relay networks and Explicit Forward Congestion Indication (EFCI) in ATM.

With MPLS LSRs, it is expected that they will implement the algorithms needed to determine an average queue size, as specified with RED and other active queue management algorithms. Thus, a congestion indication in MPLS based on the average queue size will be better matched to the ECN semantics. We thus propose that the same policies for indicating congestion be used in LSRs. The egress LSR of a label switched path (LSP) would then "fold" the congestion indication seen in the MPLS header into the forwarded IP packet. Thus, each of the LSRs also now has a means of indicating congestion to the end-systems.

3. A One-Bit ECN Field in the MPLS Header

As described in Section 2, RFC 2481 uses a two-bit ECN field for IP. The original ECN proposal in [F94] discussed both a one-bit and a two-bit implementation for ECN in the IP header. This document proposes a one-bit ECN field for the MPLS header, because of our desire to conserve space in the header, and the ability to use signaling at the time of setting up the label switched path (LSP) to overcome the need for the second bit. We propose that this bit should be one of the three experimental bits defined in [R99]. We note that by using only one of these bits, we leave two available for diff-serv drop preference, as described in [LeF99].

The ECN field has to encode three separate states: not ECT; ECT but not CE; and ECT with CE. The two-bit implementation specified in RFC 2481 implements these three separate states using two bits in the IP ECN field, the ECT bit and the CE bit. Section 9.2 of [F94] also discusses a one-bit approach where the two functions of ECT and CE are overloaded in a single bit. For this bit, one value indicates "ECT but not CE", and the other value indicates "either not ECT, or ECT with CE".

4. Role of Ingress and Egress MPLS LSRs

When the LSP is set up, the ingress and egress LSRs use signaling to indicate whether or not to participate in ECN. We envisage this as an extension to any of the existing MPLS label distribution mechanisms such as LDP or RSVP. Such signaling allows ingress and egress LSRs on an LSP to determine if all LSRs along the LSP are capable of supporting ECN. Details of these signaling mechanisms constitute work in progress and will be described in a later version of this draft.

The ingress LSR observes the IP ECN field in a packet to set the MPLS ECN field in accordance with the following table (Table 1):

	IP ECN Field	ECN-Capable MPLS LSP?	MPLS ECN Field
(1)	Not ECT	YES	Not ECT, or ECT+CE
(2)	ECT, not CE	YES	ECT, not CE
(3)	ECT, CE	YES	ECT, not CE
(4)	---	NO	Not ECT, or ECT+CE

Table 1.: Translation from the IP ECN Field to the MPLS ECN Field at Ingress.

An ingress LSR using ECN would set the MPLS ECN field to "ECT but not CE" when the IP ECN field indicates "ECT", whether or not the CE bit was set in the IP ECN field, and would have to *always* set the MPLS ECN field to "either not ECT, or ECT with CE" otherwise. Similarly, an ingress LSR not using ECN would *always* set the MPLS ECN field to "either not ECT, or ECT with CE", whether or not the IP ECN field was set to "ECT".

When the packet reaches the end of the LSP, the egress LSR now has to "fold" the MPLS ECN field back into the IP ECN header of the packet. The egress LSR knows whether the LSP is ECN-Capable or not. On the received packet, the MPLS header indicates whether congestion was experienced or not. The main row to examine in the following table (Table 2) is Row 5. It shows that the received packet has the ECN field in the MPLS header indicating congestion or that the packet flowed over a non-ECN capable LSP. However, because of the signaling at the time the LSP was set up, we know that the MPLS LSP was ECN capable. In addition, the IP ECN field was set to "ECT". Thus, the MPLS ECN field is interpreted as indicating ECT+CE (congestion was experienced in the MPLS LSP). The IP ECN field of the packet received indicates that the transport is ECN capable (ECT) and that it had not experienced congestion earlier (not CE) before entering the LSP. The IP ECN field of the forwarded packet therefore has to be set by the egress LSR to indicate congestion (CE). Thus, congestion experienced in the MPLS LSP is now successfully carried in the IP ECN field onwards to the end-system.

Row 1 of Table 2 shows that the MPLS ECN field indicates no congestion was experienced in the LSP. Thus, the IP ECN field of the packet is left unchanged. Similarly, when the original IP ECN field indicates the transport is not ECN capable or already indicates congestion was experienced, then the the IP ECN field of the packet is left unchanged.

	MPLS ECN Field	Old IP ECN Field	ECN-Capable MPLS LSP?	New IP ECN Field
(1)	ECT, not CE	---	---	Unchanged.
(2)	not ECT, or ECT+CE	Not ECT	---	Unchanged.
(3)	not ECT, or ECT+CE	ECT, CE	---	Unchanged.
(4)	not ECT, or ECT+CE	ECT, not CE	NO	Not possible.
(5)	not ECT, or ECT+CE	ECT, not CE	YES	ECT, CE

Table 2.: Translation from MPLS ECN Field back to IP ECN Field at Egress

The reason that the use of a one-bit ECN field in the MPLS header does not introduce ambiguity is that it is accompanied by the original two-bit IP ECN field, along with a prior agreement about whether the MPLS LSP was or was not ECN-Capable.

5. Role of Switches in marking ECN

	MPLS ECN Field	ECN-Capable MPLS LSP?	MPLS Congestion Action
(1)	Not ECT, or ECT+CE	No	Packet dropped.
(2)	Not ECT, or ECT+CE	Yes	Packet dropped.
(3)	ECT, not CE	Yes	MPLS ECN Field changed to "not ECT, or ECT+CE"
(4)	ECT, not CE	No	Not Possible

Table 3.: Effect of MPLS ECN Field on Congestion Action at Switch.

The congestion action taken at an LSR is based on the knowledge at the LSR of whether or not the LSP is ECN capable. This is known through signaling. If the LSP is not ECN capable, the packet is dropped as per the existing rules followed at the LSR (i.e., based on RED). If the LSP is ECN capable, and the MPLS ECN field shows that the packet is ECN-capable but has not experienced congestion upstream on the LSP, then the MPLS ECN field is changed to indicate congestion (i.e., the encoding of "Not ECT, or ECT+CE"). If the packet has indeed experienced congestion upstream, the packet is dropped. This is an inescapable consequence of the single-bit MPLS ECN field combined with internal LSRs not looking at the IP ECN field.

One functional limitation of the one-bit ECN implementation for MPLS is that once an LSR "sets" the CE state in an ECT MPLS packet, subsequent LSRs along the MPLS path cannot determine whether the packet is or is not ECN-Capable (without using the IP ECN field). Although LSRs may be able to look at the IP header (potentially with some performance impact), we do not require it. Thus, subsequent

LSRs would have to drop that packet in order to indicate congestion to the end nodes, rather than using ECN.

6. Role of Signaling to communicate ECN capability

One requirement for the one-bit ECN implementation for MPLS is that the egress LSR needs information from the ingress LSR in order to determine how to interpret the MPLS ECN Field. In particular, for a packet at the egress of the MPLS LSP whose IP ECN field indicates "ECT, not CE" and whose MPLS ECN field indicates "not ECT, or ECT+CE", the egress LSR of the MPLS LSP has to be able to determine whether to set the CE bit in the forwarded packet's IP ECN field upon decapsulation. To do this, the egress LSR has to know whether or not the ingress LSR originally set the MPLS ECN field to "ECT but not CE". This can be determined using information in the IP ECN header, assuming that the ingress LSR has previously informed the egress LSR whether it is or is not using ECN. Thus, an ingress and egress LSR would have to negotiate whether or not they are using ECN-Capability for the MPLS tunnel. Note only the ingress and egress LSRs have to examine the IP ECN header of the packet.

Based on the negotiation at the time the LSP is set up, the ingress LSR knows what the MPLS ECN field should be set to on incoming packets, especially for downstream allocation of the MPLS LSP: if the egress LSR is ECN capable and the LSRs upstream are also ECN capable, then the ingress LSR knows to initially set the MPLS ECN field to "ECT but not CE" when the IP ECN field indicated "ECT". If the LSRs in the MPLS LSP are not ECN capable, then the ingress LSR always sets the MPLS ECN bit to "not ECT or ECT+CE". The egress LSR, knowing that the LSP is ECN capable through signaling, folds the MPLS ECN field into the outgoing IP ECN field according to Table 2.

7. Summary

We have proposed that MPLS LSRs exploit the capability provided in Explicit Congestion Notification (ECN) to provide an early indication of congestion without depending solely on packet dropping as a means of congestion indication. Given that MPLS LSRs are likely to use some form of active queue management, the use of ECN potentially improves the service obtained in an MPLS LSP with respect to packet loss and end-end delay. The proposal requires the use of a single bit in the MPLS header for indicating congestion, and signaling while setting up the LSP. The signaling provides the indication to the ingress and egress LSRs the action they need to take with respect to ECN: the initialization of the MPLS ECN field at the ingress and the folding of the MPLS ECN indication into the forwarded IP packet's ECN field at the egress.

8. References

[ECN] The ECN Web Page, URL "http://www.aciri.org/floyd/ecn.html".

[F94] Floyd, S., "TCP and Explicit Congestion Notification", ACM Computer Communication Review, V. 24 N. 5, October 1994, p. 10-23. URL "ftp://ftp.ee.lbl.gov/papers/tcp_ecn.4.ps.Z".

[FBR99] Floyd, Black, and Ramakrishnan, IPsec Interactions with ECN, internet-draft draft-ipsec-ecn-00.txt, April 1999. URL "http://www.ietf.cnri.reston.va.us/internet-drafts/draft-ipsec-ecn-00.txt".

[LeF99] Le Faucheur, F., et al., "MPLS Support of Differentiated Services over PPP links", internet draft, draft-lefaucheur-mpls-diff-ppp-00.txt, June 1999. URL "http://www.ietf.cnri.reston.va.us/internet-drafts/draft-lefaucheur-mpls-diff-ppp-00.txt"

[RFC2309] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J. and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, April 1998.

[RFC2481] K. K. Ramakrishnan and Sally Floyd, "A Proposal to add Explicit Congestion Notification (ECN) to IP", RFC 2481, January 1999. URL "ftp://ftp.isi.edu/in-notes/rfc2481.txt".

[R99] Rosen, E., et al., "MPLS Label Stack Encoding", internet draft, draft-ietf-mpls-label-encaps-04.txt, April 1999. URL "http://www.ietf.cnri.reston.va.us/internet-drafts/draft-ietf-mpls-label-encaps-04.txt".

9. Security Considerations

Security considerations are equivalent to those for normal IP ECN and ECN with IPsec, which are discussed extensively in [FBR99].

AUTHORS' ADDRESSES

Sally Floyd
AT&T Center for Internet Research at ICSI (ACIRI)
Phone: +1 (510) 642-4274 x189
Email: floyd@aciri.org
URL: http://www-nrg.ee.lbl.gov/floyd/

K. K. Ramakrishnan
AT&T Labs. Research
Phone: +1 (973) 360-8766
Email: kkrama@research.att.com
URL: http://www.research.att.com/info/kkrama

Bruce Davie
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824
E-mail: bsd@cisco.com

This draft was created in June 1999.
It expires December 1999.