Internet Engineering Task Force S. Floyd
INTERNET-DRAFT ICIR
draft-floyd-ecn-alternates-00.ps 21 April 2005
Expires: October 2005

Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field

Status of this Memo

By submitting this Internet-Draft, we certify that any applicable patent or other IPR claims of which we are aware have been disclosed, or will be disclosed, and any of which we become aware will be disclosed, in accordance with RFC 3668 (BCP 79).

By submitting this Internet-Draft, we accept the provisions of Section 3 of RFC 3667 (BCP 78).

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

There have been a number of proposals for alternate semantics for the ECN field in the IP header [ECN]. This document discusses some of the issues in defining alternate semantics for the ECN field, and specifies requirements for a safe co-existence in an Internet that could include routers that do not understand the defined alternate semantics. This document evolved as a result of discussions with the authors of one recent proposal for such alternate semantics.

Table of Contents

1. Introduction 2
2. An Overview of the Issues 2
3. Signalling the Use of Alternate ECN Semantics 3
3.1. Using the Diffserv Field for Signalling 3
4. Issues of Incremental Deployment 4
4.1. Option 1: Unsafe for Deployment in the Internet 5
4.2. Option 2: Verification that Routers Understand the Alternate Semantics 6

4.3. Option 3: Friendly Co-existence with Competing Traffic 6
5. Evaluation of the Alternate-ECN Semantics 7
5.1. Verification of Feedback from the Router 7
5.2. Co-existence with Competing Traffic 8
5.3. A General Evaluation of the Alternate-ECN Semantics 8
6. Who Wants to Use Alternate Semantics for the ECN Codepoint? 8
7. Security Considerations 8
8. Acknowledgements 8
9. Conclusions 9
10. Normative References 9
11. Informative References 9
IANA Considerations 9
AUTHORS' ADDRESSES 10
Full Copyright Statement 10
Intellectual Property 10

,

1. Introduction

RFC 3168, a Proposed Standard document, defines the ECN field in the IP header, and specifies the semantics for the codepoints for the ECN field. However, end nodes could specify the use of alternate semantics for the ECN field, e.g., using codepoints in the diffserv field of the IP header. This document describes some of the issues that arise in specifying such alternate semantics for the ECN field, and gives requirements for a safe co-existence in a world using the default ECN semantics (or using no ECN at all).

2. An Overview of the Issues

In this section we discuss some of the issues that arise if some of the traffic in a network consists of alternate-ECN traffic (i.e., traffic using alternate semantics for the ECN field).

The first issue concerns how routers know which ECN semantics to use with which packets in the network:

(1) How does the connection indicate to the router that its packets are using alternate-ECN semantics? Is the specification of alternate-ECN semantics robust and unambiguous? If not, is this a problem?

As an example, in most of the proposals for alternate-ECN semantics, a diffserv field is used to specify the use of alternate-ECN semantics. Do all routers that understand this diffserv codepoint understand that it uses alternate-ECN semantics, or not? Can the diffserv field be re-written in the middle of the network? If so, is this a problem?

This is discussed in more detail in Section 3 below.

A second issue is that of incremental deployment in a network where some routers only use the default ECN semantics, and other routers might not use ECN at all. In this document we use the phrase "new routers" to refer to the routers that understand the alternate-ECN

semantics, and "old routers" to refer to routers that don't understand or aren't willing to use the alternate-ECN semantics. The possible existence of old routers raises several questions:

(2) How does the possible presence of old routers affect the performance of the alternate-ECN connections?

(3) How does the possible presence of old routers affect the co-existence of the alternate-ECN traffic with competing traffic on the path?

These issues are discussed in Section 4 below.

A final issue is that of the general evaluation of the alternate-ECN semantics:

(4) How well does the alternate-ECN traffic perform, and how well does it co-exist with competing traffic on the path, in a "clean" environment with new routers and with the unambiguous specification of the use of alternate-ECN semantics?

These issues are discussed in Section 5 below.

3. Signalling the Use of Alternate ECN Semantics

This section discusses question (1) from Section 2:

(1) How does the connection indicate to the router that its packets are using alternate-ECN semantics? Is the specification of alternate-ECN semantics robust and unambiguous? If not, is this a problem?

The assumption of this document is that when alternate semantics are defined for the ECN field, a codepoint in the diffserv field is used to signal the use of these alternate ECN semantics to the router. That is, the end host sets the codepoint in the diffserv field to indicate to routers that alternate semantics to the ECN field are being used. Routers that understand this diffserv codepoint would know to use the alternate semantics for interpreting and setting the ECN field. Old ECN-capable routers that do not understand this diffserv codepoint would use the default ECN semantics in interpreting and setting the ECN field.

We note that this does not exclude the possibility of using other methods, including out-of-band mechanisms, for signalling the use of alternate semantics for the ECN field. The considerations in the rest of this document apply regardless of the method used to signal the use of alternate semantics for the ECN field.

3.1. Using the Diffserv Field for Signalling

We note that the default ECN semantics defined in RFC 3168 are the current default semantics for the ECN field, regardless of the contents of any other fields in the IP header. In particular, the default ECN semantics don't just apply for best-effort traffic with a codepoint of '000000' for the diffserv field - the default ECN semantics currently apply

regardless of the contents of the diffserv field.

There are two ways to use the diffserv field to signal the use of alternate ECN semantics. One way is to use an existing diffserv codepoint, and to modify the current definition of that codepoint, through approved IETF processes, to specify the use of alternate ECN semantics with that codepoint. A second way is to define a new diffserv codepoint, and to specify the use of alternate ECN semantics with that codepoint. We note that the first of these two mechanisms raises the possibility that some routers along the path will understand the diffserv codepoint but will use the default ECN semantics with this diffserv codepoint, or won't use ECN at all, and that other routers will use the alternate ECN semantics with this diffserv codepoint.

4. Issues of Incremental Deployment

This section discusses questions (2) and (3) posed in Section 2:

(2) How does the possible presence of old routers affect the performance of the alternate-ECN connections?

(3) How does the possible presence of old routers affect the co-existence of the alternate-ECN traffic with competing traffic on the path?

When alternate semantics are defined for the ECN field, it is necessary to ensure that there are no problems caused by old routers along the path that don't understand the alternate ECN semantics.

One possible problem is that of poor performance for the alternate-ECN traffic. Is it essential to the performance of the alternate-ECN traffic that all routers along the path understand the alternate-ECN semantics? If not, what are the possible consequences, for the alternate-ECN traffic itself, when some old routers along the path don't understand the alternate-ECN semantics? These issues have to be answered in the context of each specific proposal for alternate ECN semantics.

A second specific problem is that of possible unfair competition with other traffic along the path. If there is an old router along the path that doesn't use ECN, that old router could drop packets from the alternate-ECN traffic, and expect the alternate-ECN traffic to reduce its sending rate as a result. Does the alternate-ECN traffic respond to packet drops as an indication of congestion?

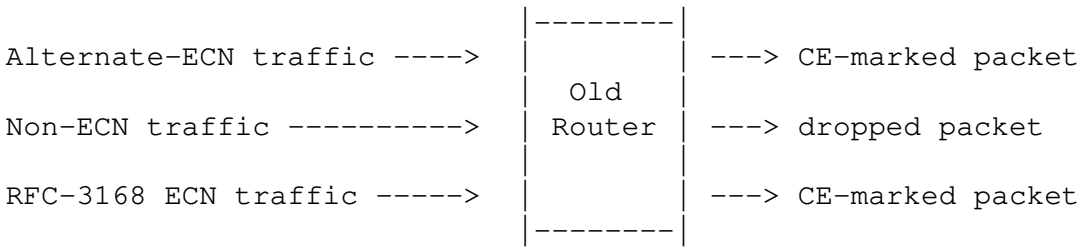

Figure 1: Alternate-ECN traffic with an old router using RFC-3168 ECN.

Similarly, what if there is an old router along the path that understands only the default ECN semantics from RFC-3168, as in Figure 1 above? In times of congestion, the old default-ECN router could see an alternate-ECN packet with one of the ECN-Capable Transport (ECT) codepoints set in the ECN field in the IP header, as defined in RFC 3168, and set the Congestion Experienced (CE) codepoint in the ECN field as an alternative to dropping the packet. The router in this case would expect the alternate-ECN connection to respond, in terms of congestion control, as it would if the packet has been dropped. If the alternate-ECN traffic fails to respond appropriately to the CE codepoint being set by an old router, this could increase the aggregate traffic arriving at the old router, resulting in an increase in the packet-marking and packet-dropping rates at that router, further resulting in the alternate-ECN traffic crowding out the other traffic competing for bandwidth on that link.

Basically, there are three possibilities for avoiding scenarios where the presence of old routers along the path results in the alternate-ECN traffic competing unfairly with other traffic along the path:

Option 1: Alternate-ECN traffic is clearly understood as unsafe for deployment in the global Internet; or

Option 2: All alternate-ECN traffic deploys some mechanism for verifying that all routers on the path understand and agree to use the alternate ECN semantics for this traffic; or

Option 3: The alternate-ECN semantics are defined in such a way as to ensure the fair and peaceful co-existence of the alternate-ECN traffic with best-effort and other traffic, even in environments that include old routers that do not understand the alternate-ECN semantics.

Each of these alternatives is explored in more detail below.

4.1. Option 1: Unsafe for Deployment in the Internet

The first option specified above is for the alternate-ECN traffic to be clearly understood as only suitable for enclosed environments, and as unsafe for deployment in the global Internet. This would mean that it would be unsafe for packets using the alternate ECN semantics to be unleashed in the global Internet, in order to avoid the chance of the alternate-ECN traffic traversing an old router that don't understand the alternate semantics. This document doesn't comment on whether a mechanism would be required to ensure that the alternate-ECN semantics would not be let loose on the global Internet. This document

also doesn't comment on the chances that this scenario would be considered acceptable for standardization by the IETF community.

4.2. Option 2: Verification that Routers Understand the Alternate Semantics

The second option specified above is for the alternate-ECN traffic to include a mechanism for ensuring that all routers along the path understand and agree to the use of the alternate ECN semantics for this traffic. As an example, such a mechanism could consist of a field in an IP option that all routers along the path decrement if they agree to use the alternate ECN semantics with this traffic. (A similar mechanism is proposed for Quick-Start, for verifying that all of the routers along the path understand the Quick-Start IP Option [QuickStart].) Using such a mechanism, a sender could have reasonable assurance that the packets that are sent specifying the use of alternate ECN semantics only traverse routers that in fact understand and agree to use these alternate semantics for these packets.

4.3. Option 3: Friendly Co-existence with Competing Traffic

The third option specified above is for the alternate ECN semantics to be defined so that traffic using the alternate semantics would co-exist safely in the Internet on a path with one or more old routers that use only the default ECN semantics. In this scenario, a connection sending alternate-ECN traffic would have to respond appropriately to a CE packet (a packet with the ECN codepoint "11") received at the receiver, using a conformant congestion control response. Hopefully, the connection sending alternate-ECN traffic would also respond appropriately to a dropped packet, that could be a congestion indication from a router that doesn't use ECN at all.

RFC 3168 defines the default ECN semantics as follows:

"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. For example, for ECN-Capable TCP the source TCP is required to halve its congestion window for any window of data containing either a packet drop or an ECN indication."

The only conformant congestion control mechanisms currently standardized in the IETF are TCP [RFC2581] and protocols using TCP-like congestion control (e.g., SCTP [RFC2960], DCCP with CCID-2 [DCCP]), and TCP-Friendly Rate Control (TFRC) and protocols with TFRC-like congestion control (e.g., DCCP using CCID-3 [DCCP]). TCP uses Additive-Increase Multiplicative-Decrease congestion control, and responds to the loss or ECN-marking of a single packet by halving its congestion window. In contrast, the equation-based congestion control mechanism in TFRC estimates the loss event rate over some period of time, and uses a sending rate that would be comparable, in packets per round-trip-time, to that of a TCP connection experiencing the same loss event rate.

So what are the requirements in order for alternate-ECN traffic to compete appropriately with other traffic on a path through an old router that doesn't understand the alternate ECN semantics (and therefore might be using the default ECN semantics)?

The first requirement is that if a packet is marked with the ECN codepoint "11" in the network, this marking is not changed on the packet's way to the receiver (unless the packet is dropped before it reaches the receiver). This requirement is necessary to ensure that congestion indications from a default-ECN router make it to the transport receiver.

The second requirement is that the end-nodes respond in a TCP-friendly manner to packets received that are marked with the ECN codepoint "11" (because these packets might have come from a congested router that understands only the default ECN semantics). This would ensure that the traffic using the alternate semantics does not 'bully' competing traffic that it might encounter along the path, and does not drive up congestion on the shared link inappropriately.

As discussed earlier, the current conformant congestion control responses to a dropped or default-ECN-marked packet consist of TCP and TCP-like congestion control, and of TFRC (TCP-Friendly Rate Control). Another possible response considered in RFC 3714, but not standardized in a standards-track document, is that of simply terminating an alternate-ECN connection for a period of time if the long-term sending rate is higher that would be that of a TCP connection experiencing the same packet dropping or marking rates [RFC3714]. We note that the use of such a congestion control response to CE-marked packets would require specification of time constants for measuring the loss rates and for stopping transmission, and would require a consideration of issues of packet size.

5. Evaluation of the Alternate-ECN Semantics

This section discusses question (4) posed in Section 2:

(4) How well does the alternate-ECN traffic perform, and how well does it co-exist with competing traffic on the path, in a "clean" environment with new routers and with the unambiguous specification of the use of alternate-ECN semantics?

5.1. Verification of Feedback from the Router

One issue in evaluating the alternate-ECN semantics concerns mechanisms to discourage lying from the transport receiver to the transport sender. In many cases the sender is a server that has an interest in using the alternate-ECN semantics correctly, while the receiver has more incentives for lying about the congestion experienced along the path.

In the default ECN semantics, two of the four ECN codepoints are used for ECN-Capable(0) and ECN-Capable(1). The use of two codepoints for ECN-Capable, instead of one, permits the data sender to verify receiver's reports that packets were actually received unmarked at the receiver. In particular, the sender can specify that the receiver report to the sender whether each unmarked packet was received ECN-Capable(0) or ECN-Capable(1), as discussed in RFC 3540 [RFC 3540].

If alternate semantics for the ECN codepoint don't include the use of two separate codepoints to indicate ECN-Capable, then the connections using those semantics have lost the ability to verify that the data receiver is accurately reporting the received ECN codepoint to the data sender. In this case, it might be necessary for the alternate-ECN

framework to include alternate mechanisms for ensuring that the data receiver is reporting feedback appropriately to the sender.

5.2. Co-existence with Competing Traffic

A second general issue concerns the co-existence of alternate-ECN traffic with competing traffic along the path, in a clean environment where all routers understand and are willing to use the alternate-ECN semantics for the traffic that specifies its use.

If the traffic using the alternate-ECN semantics is best-effort traffic, then it is subject to the general requirement of fair competition with TCP and other traffic along the path [RFC2914].

If the traffic using the alternate-ECN semantics is diffserv traffic, then the requirements are governed by the overall guidelines for that class of diffserv traffic. It is beyond the scope of this document to specify the requirements, if any, for the co-existence of diffserv traffic with other traffic on the link; this should be addressed in the specification of the diffserv codepoint itself.

5.3. A General Evaluation of the Alternate-ECN Semantics

A third general issue concerns the evaluation of the general merits of the proposed alternate-ECN semantics. Again, it would be beyond the scope of this document to specify requirements for the general evaluation of alternate-ECN semantics.

6. Who Wants to Use Alternate Semantics for the ECN Codepoint?

There have been a number of proposals in the IETF and in the research community for alternate semantics for the ECN codepoint [ECN]. One such proposal, [BCF05], proposes an alternate-ECN semantics for real-time inelastic traffic such as voice, video conferencing, and multimedia streaming in DiffServ networks. In this proposal, the alternate-ECN semantics would provide information about two levels of congestion experienced along the path, where "it is left up to the application designers to define how end-systems should react to ECN bit marking" [BCF05]. Some of the other proposals for alternate ECN semantics are listed on the ECN Web Page [ECN].

7. Security Considerations

This document doesn't propose any new mechanisms for the Internet protocol, and therefore doesn't introduce any new security considerations.

8. Acknowledgements

This document is based in part on conversations with Jozef Babiarz, Kwok Ho Chan, and Victor Firoiu on their proposal for an alternate use of the ECN field in DiffServ

environments. Many thanks to Francois Le Faucheur for feedback recommending that the document include a section at the beginning discussing the potential issues that need to be addressed. Thanks also to Fred Baker, David Black, and to members of the TSVWG working group for feedback on these issues.

9. Conclusions

This document has discussed some of the issues to be considered in the specification of alternate semantics for the ECN field in the IP header.

10. Normative References

11. Informative References

[BCF05] J. Babiarz, K. Chan, and V. Firoiu, Congestion Notification Process for Real-Time Traffic, internet-draft draft-babiarz-tsvwg-rtecn-03, work in progress, February 2005.

[DCCP] DCCP Web Page, URL "http://www.icir.org/kohler/dccp/".

[ECN] ECN Web Page, URL "www.icir.org/floyd/ecn.html".

[QuickStart] Quick-Start Web Page, URL "http://www.icir.org/floyd/quickstart.html".

[RFC2581] M. Allman, V. Paxson, and W. Stevens, TCP Congestion Control, RFC 2581, Proposed Standard, April 1999.

[RFC2914] S. Floyd, Congestion Control Principles, RFC 2914, Best Current Practice, September 2000.

[RFC2960] R. Stewart et al, Stream Control Transmission Protocol, RFC 2960, October 2000.

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

[RFC3540] N. Spring, D. Wetherall, and D. Ely, Robust Explicit Congestion Notification (ECN) Signaling with Nonces, RFC 3540, Experimental, June 2003.

[RFC3714] S. Floyd and J. Kempf, Editors, IAB Concerns Regarding Congestion Control for Voice Traffic in the Internet, RFC 3714, Informational, March 2004.

IANA Considerations

There are no IANA considerations in this document.

AUTHORS' ADDRESSES

Sally Floyd
Phone: +1 (510) 666-2989
ICIR (ICSI Center for Internet Research)
Email: floyd@icir.org
URL: http://www.icir.org/floyd/

Full Copyright Statement

Copyright (C) The Internet Society 2004. This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.