Network Working Group                                          S. Floyd
Request for Comments: 4774                                         ICIR
BCP: 124                                                  November 2006
Category: Best Current Practice

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

## Status of This Memo

This document specifies an Internet Best Current Practices for the Internet Community, and requests discussion and suggestions for improvements. Distribution of this memo is unlimited.

## Copyright Notice



## Abstract

There have been a number of proposals for alternate semantics for the Explicit Congestion Notification (ECN) field in the IP header RFC 3168. This document 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. This document evolved as a result of discussions with the authors of one recent proposal for such alternate semantics.

Table of Contents



## 1. Introduction

[RFC3168], 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.

There have been a number of proposals in the IETF and in the research community for alternate semantics for the ECN codepoint. One such proposal, [BCF05], proposes 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 [BCF05]. Another research proposal, [XSSK05], proposes a low-complexity protocol, Variable-structure congestion Control Protocol (VCP), that uses the two bits in the ECN field to indicate low-load, high-load, and overload (congestion), where transport protocols can increase more rapidly during the low-load regime. Some of the proposals for alternate ECN semantics are for when ECN is used in an edge-to-edge context between gateways at the edge of a network region, e.g., for pre-congestion notification for admissions control [BESFC06]. Other proposals for alternate ECN semantics are listed on the ECN Web Page [ECN].

The definition of multiple semantics for the ECN field could have significant implications on both host and router implementations. There is a huge base of installed hosts and routers in the Internet, and in other IP networks, and updating these is an enormous and potentially expensive undertaking. Some existing devices might be able to support the new ECN semantics with only a software upgrade and without significant degradation in performance. Some other equipment might be able to support the new semantics, but with a degradation in performance -- which could range from trivial to catastrophic. Some other deployed equipment might be able to support the new ECN semantics only with a hardware upgrade, which, in some cases, could be prohibitively expensive to deploy on a very wide scale. For these reasons, it would be difficult and would take a significant amount of time to universally deploy any new ECN semantics. In particular, routers can be difficult to upgrade, since small routers sometimes are not updated frequently, and large routers commonly have specialized forwarding paths to facilitate high performance.

This document describes some of the technical issues that arise in specifying alternate semantics for the ECN field, and gives requirements for a safe coexistence 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 issues include the following: (1) how routers know which ECN semantics to use with which packets; (2) incremental deployment in a network where some routers use only the default ECN semantics or do not use ECN at all; (3) coexistence of alternate-ECN traffic with competing traffic on the path; and (4) a general evaluation of the alternate ECN semantics.

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

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? Diffserv allows routers to re-mark DiffServ Code Point (DSCP) values within the network; what is the effect of this on the alternate ECN semantics?

This is discussed in more detail in Section 3 below.

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

(3) The possible existence of old routers also raises the question of how the presence of old routers affects the coexistence of the alternate-ECN traffic with competing traffic on the path.

Issues (2) and (3) are discussed in Section 4 below.

(4) A final issue is that of the general evaluation of the alternate ECN semantics:

How well does the alternate-ECN traffic perform, and how well does it coexist 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.

## 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.

In general, the diffserv codepoints are used to signal the per-hop behavior at router queues. One possibility would be to use one diffserv codepoint to signal a per-hop behavior with the default ECN semantics, and a separate diffserv codepoint to signal a similar per-hop behavior with the alternate ECN semantics. Another possibility would be to use a diffserv codepoint to signal the use of best-effort per-hop queueing and scheduling behavior, but with alternate ECN semantics. A detailed discussion of these issues is beyond the scope of this document.

We note that this discussion 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 apply for more than 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 coexistence 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?

```
                                  |--------|
  Alternate-ECN traffic ---->     |        |  ---> CE-marked packet
                                  |   Old  |
  Non-ECN traffic ---------->     | Router |  ---> dropped packet
                                  |        |
  RFC-3168 ECN traffic ----->     |        |  ---> CE-marked packet
                                  |--------|
```

Figure 1: Alternate-ECN traffic, an old router, using RFC-3168 ECN, that is congested and ready to drop or mark the arriving packet.

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 coexistence 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. Specifically, this would mean that it would be unsafe for packets using the alternate ECN semantics to be unleashed in the global Internet. This restriction would prevent the alternate-ECN traffic from traversing an old router outside of the enclosed environment that didn'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. Note, however, that most existing routers are optimized for IP packets with no options, or with only some very well-known and simple IP options. Thus, the definition and use of any new IP option may have a serious detrimental effect on the performance of many existing IP routers.

Such a mechanism should be robust in the presence of paths with multi-path routing, and in the presence of routing or configuration changes along the path while the connection is in use. In particular, if this option is used, connections could include some form of monitoring for changes in path behavior and/or periodic monitoring that all routers along the path continue to understand the alternate ECN semantics.

### 4.3. Option 3: Friendly Coexistence 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 coexist 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, which could be a congestion indication from a router that doesn't use ECN.

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 ([RFC4340], [RFC4341])), and TCP-Friendly Rate Control (TFRC) [RFC3448], and protocols with TFRC-like congestion control (e.g., DCCP using CCID-3 [RFC4342]). 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 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 and second requirements below concern compatibility between traffic using alternate ECN semantics and routers using default ECN semantics.

The first requirement for compatibility with routers using default ECN 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.

A second requirement for compatibility with routers using default ECN is that the end-nodes respond to packets that are marked with the ECN codepoint "11" in a way that is friendly to flows using IETF-conformant congestion control. This requirement is needed because the "11"-marked packets might have come from a congested router that understands only the default ECN semantics, and that expects that end-nodes will respond appropriately to CE packets. This requirement would ensure that the traffic using the alternate semantics does not 'bully' competing traffic that it might encounter along the path, and that it does not drive up congestion on the shared link inappropriately.

Additional requirements concern compatibility between traffic using default ECN semantics and routers using alternate ECN semantics. This situation could occur if a diffserv codepoint using default ECN semantics is redefined to use alternate ECN semantics, and traffic from an "old" source traverses a "new" router. If the router "knows" that a packet is from a sender using alternate semantics (e.g., because the packet is using a certain diffserv codepoint, and all packets with that diffserv codepoint use alternate semantics for the ECN field), then the requirements below are not necessary, and the rules for the alternate semantics apply.

A requirement for compatibility with end-nodes using default ECN is that if a packet that *could* be using default semantics is marked with the ECN codepoint "00", this marking must not be changed to "01", "10", or "11" in the network. This prevents the packet from being represented incorrectly to a default-ECN router downstream as ECN-Capable. Similarly, if a packet that *could* be using default semantics is marked with the ECN codepoint "01", then this codepoint should not be changed to "10" in the network (and a "10" codepoint should not be changed to "01"). This requirement is necessary to avoid interference with the transport protocol's use of the ECN nonce [RFC3540].

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 than 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 coexist 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 incentive to lie 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 the 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 [RFC3540]. This use of ECN-Capable(0) and ECN-Capable(1) is independent of the semantics of the other ECN codepoints, and could be used, if desired, with alternate semantics for the other codepoints.

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. As one possibility, policers could be used in routers to ensure that end nodes are responding appropriately to marked packets.

### 5.2. Coexistence with Competing Traffic

A second general issue concerns the coexistence 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 coexistence of diffserv traffic with other traffic on the link; this should be addressed in the specification of the diffserv codepoint itself.

## 5.3. Proposals for Alternate ECN with Edge-to-Edge Semantics

RFC 3168 specifies the use of the default ECN semantics by an end-to-end transport protocol, with the requirement 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" ([RFC3168], Section 5). In contrast, some of the proposals for alternate ECN semantics are for ECN used in an edge-to-edge context between gateways at the edge of a network region, e.g., [BESFC06].

When alternate ECN is defined with edge-to-edge semantics, this definition needs to ensure that the edge-to-edge semantics do not conflict with a connection using other ECN semantics end-to-end. One way to avoid conflict would be for the edge-to-edge ECN proposal to include some mechanism to ensure that the edge-to-edge ECN is not used for connections that are using other ECN semantics (standard or otherwise) end-to-end. Alternately, the edge-to-edge semantics could be defined so that they do not conflict with a connection using other ECN semantics end-to-end.

## 5.4. Encapsulated Packets

RFC 3168 has an extensive discussion of the interactions between ECN and IP tunnels, including IPsec and IP in IP. Proposals for alternate ECN semantics might interact with IP tunnels differently than default ECN. As a result, proposals for alternate ECN semantics must explicitly consider the issue of interactions with IP tunnels.

## 5.5. 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. Security Considerations

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

## 7. 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.

Specifications of alternate ECN semantics must clearly state how they address the issues raised in this document, particularly the issues discussed in Section 2. In addition, specifications for alternate ECN semantics must meet the requirements in Section 5.2 for coexistence with competing traffic.

## 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 Mark Allman, Fred Baker, David Black, Gorry Fairhurst, and to members of the TSVWG working group for feedback on these issues.

## 9. Normative References

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

## 10. Informative References

[BCF05] Babiarz, J., Chan, K., and V. Firoiu, "Congestion Notification Process for Real-Time Traffic", Work in Progress, July 2005.

[BESFC06] Briscoe, B., et al., "An edge-to-edge Deployment Model for Pre-Congestion Notification: Admission Control over a DiffServ Region", Work in Progress, June 2006.

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

[QuickStart] S. Floyd, M. Allman, A. Jain, and P. Sarolahti, "Quick-Start for TCP and IP", Work in Progress, October 2006.

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

[RFC2914] Floyd, S., "Congestion Control Principles", BCP 41, RFC 2914, September 2000.

[RFC2960] Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L., and V. Paxson, "Stream Control Transmission Protocol", RFC 2960, October 2000.

[RFC3448] Handley, M., Floyd, S., Padhye, J., and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification", RFC 3448, January 2003.

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

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

[RFC4340] Kohler, E., Handley, M., and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", RFC 4340, March 2006.

[RFC4341] Floyd, S. and E. Kohler, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control", RFC 4341, March 2006.

[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, March 2006.

[XSSK05] Y. Xia, L. Subramanian, I. Stoica, and S. Kalyanaraman, One More Bit Is Enough, SIGCOMM 2005, September 2005.

Author's Address

Sally Floyd
ICIR (ICSI Center for Internet Research)

Phone: +1 (510) 666-2989
EMail: floyd@icir.org
URL: http://www.icir.org/floyd/

## Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.