

Internet Draft

Francois Le Faucheur  
Anna Charny  
Cisco Systems, Inc.Bob Briscoe  
Phil Eardlwy  
BTJoe Barbiaz  
Kwok-Ho Chan  
Nortel[draft-lefaucheur-rsvp-ecn-00.txt](#)

Expires: April 2006

October 2005

RSVP Extensions for Admission Control  
over Diffserv using Pre-congestion Notification

## Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

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/lid-abstracts.txt>.

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

## Abstract

This document specifies the extensions to RSVP for support of the Controlled Load (CL) service over a Diffserv cloud using Pre-Congestion Notification as defined in [\[CL-ARCH\]](#).

#### Copyright Notice

Copyright (C) The Internet Society. (2005)

#### Specification of Requirements

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

### 1. Introduction

[RSVP] defines the Resource reSerVation Protocol which can be used by applications to request resources from the network. The network responds by explicitly admitting or rejecting these RSVP requests. Certain applications that have quantifiable resource requirements express these requirements using Intserv parameters as defined in the appropriate Intserv service specifications ([GUARANTEED], [CONTROLLED]).

[CL-ARCH] describes a framework to achieve a Controlled Load (CL) service by using distributed measurement-based admission control edge-to-edge, i.e. within a particular region of the Internet. The measurement made is of CL packets that have their Congestion Experienced (CE) codepoint set as they travel across the edge-to-edge region. Setting the CE codepoint, which is under the control of a new Pre-congestion Marking behaviour, provides an "early warning" of potential congestion. This information is used by the ingress node of the edge-to-edge region to decide whether to admit a new CL microflow.

[CL-ARCH] also describes how the framework uses rate-based pre-emption to maintain the CL service to as many admitted microflows as possible even after localised failure and routing changes in the interior of the edge-to-edge region.

The edge-to-edge architecture of [\[CL-ARCH\]](#) is a building block in delivering an end-to-end CL service. The approach is similar to that described in [INTSERV-DIFFERV] for Integrated services operation over Diffserv networks. Like [INTSERV-DIFFERV], an IntServ class (CL in our case) is achieved end-to-end, with a CL-region viewed as a single reservation hop in the total end-to-end path. Interior nodes of the CL-region do not process flow signalling nor do they hold state. [CL-ARCH] assumes that the end-to-end signalling mechanism is RSVP.

This document specifies the extensions to RSVP for support of the Controlled Load (CL) service over a Diffserv cloud using Pre-Congestion Notification as defined in [\[CL-ARCH\]](#).

[1.1.](#) Changes from previous versions

- This is the initial version of the document

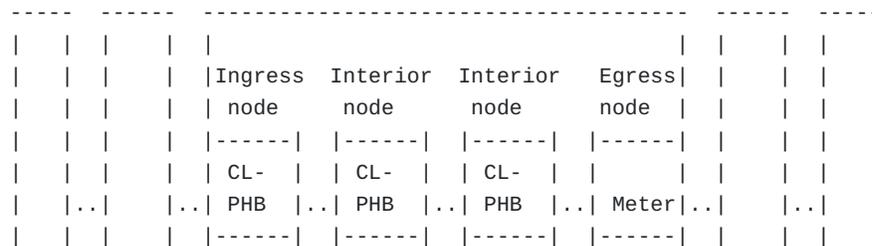
[2.](#) Definitions

For readability, a number of definitions from [\[CL-ARCH\]](#) are repeated here:

- o Ingress Edge: a node which is an ingress gateway to the CL-region. A CL-region may have several ingress nodes.
- o Egress Edge: a node which is an egress gateway from the CL-region. A CL-region may have several egress nodes.
- o Interior node: a node which is part of the CL-region, but isn't an Ingress or Egress Edge.
- o CL-region: A region of the Internet in which all nodes run the Pre-Congestion Notification scheme defined in [\[CL-MARKING\]](#) and where all traffic enters/leaves through an Ingress/Egress Edge. A CL-region is a DiffServ region (either a single DiffServ domain or set of contiguous DiffServ domains).
- o Congestion-Level-Estimate: the bits in CL packets that have the Congestion Experienced (CE) codepoint set, divided by the bits in all CL packets. It is calculated as an exponentially weighted moving average. It is calculated by an egress node for CL packets from a particular ingress node.

[3.](#) Overview of RSVP extensions and Operations

[3.1.](#) Reference Model



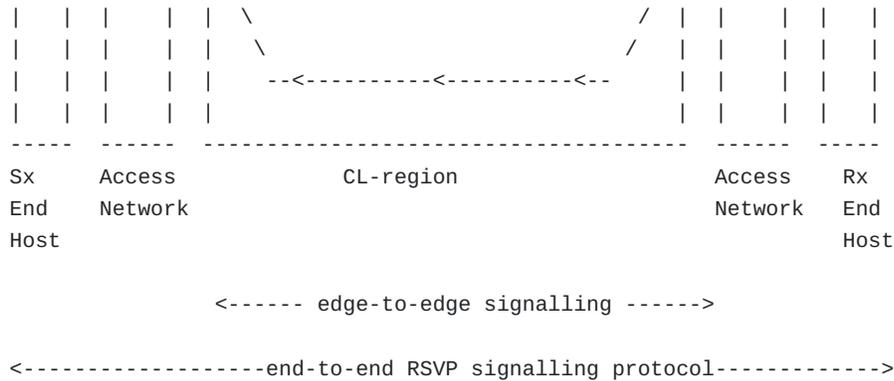


Figure 1: Reference Model

### 3.2. Overview of Procedures for Admission Control of New Reservations

As mentioned earlier, [CL-ARCH] describes a framework to achieve a Controlled Load (CL) service by using distributed measurement-based admission control edge-to-edge, i.e. within a particular region of the Internet. This section describes RSVP operations to support such an admission control scheme relying on Pre-Congestion Notification in the edge-to-edge region.

When a new Path message is received by Ingress Edge, the Ingress Edge does regular RSVP processing and forwards the Path towards destination.

All the ECN-capable Interior nodes are not RSVP-capable and thus simply ignore the Path message.

When the Path message arrives at the Egress Edge, the Egress Edge processes it as per regular RSVP processing augmented with the following rules:

- 1) The Egress Edge does NOT perform the RSVP-TTL vs IP TTL-check and does NOT update the ADspec Break bit. This is because the whole CL-region is effectively handled by RSVP as a virtual "link" on which Integrated Service is indeed supported (and admission control performed) so that the Break bit MUST not be set.
- 2) As an option, the Egress Edge may check, at the time of initial Path processing, whether it has a valid value for the corresponding Congestion-Level-Estimate and if not it may send a PathErr message to the Ingress Edge with "CL-ECN Probes Required" Error Code. This minimizes call set up time as it allows probes to be generated by the Ingress Edge and measured by the Egress Edge while the Path is traveling towards the receiver and while the Resv travels back from the receiver.

Then the Ingress Edge forwards the Path message towards the receiver.

[Editor Note: discussion on Adspec update to be added]

When the Resv message is received by the Egress Edge (from the downstream side), the Egress Edge performs regular RSVP processing (including performing admission control for the segment downstream of the Egress Edge) augmented with the procedures described in this section.

The Egress Edge MUST include the new CL-ECN object in the Resv message transmitted to the RSVP P\_HOP (which is the Ingress Edge). The CL-ECN object MUST convey the current Pre-Congestion Notification Congestion-Level-Estimate as measured by the Egress Edge from the corresponding Ingress Edge to itself. Details for computing the Congestion-Level-estimate can be found in [[CL-ARCH](#)] and [[CL-MARKING](#)].

If the Egress Edge does not have a current value for the Congestion-Level-estimate for the corresponding Ingress Edge (because there was no traffic received by the Egress Edge from that Ingress Edge) and it has not already requested the Ingress Edge to generate probes, the Egress Edge:

- 1) triggers a timer and puts the Resv message processing on hold
- 2) sends a PathErr message towards the Ingress Edge with the new Error Code of "CL-ECN Probes Required" specified in this document, in order to instruct the Ingress Edge to generate the necessary probe traffic to enable the Egress Edge to compute the Congestion-Level-Estimate from that Ingress Edge
- 3) When timer expires the Resv processing resumes. Assuming the Congestion-Level-Estimate is now available, the Egress Edge can include it in the CL-ECN object and complete Resv processing. If the Congestion-Level-Estimate is still available, the Egress Edge may loop again a few times through step 1) and 2). After a given number of times, the Egress Edge MUST send a ResvErr towards the receiver with ErrorCode "Admission Control Failure"

[Editor note: approach in previous paragraph may be revisited to try avoid having to "put Resv message processing on hold".]

The Egress Edge will then forward the Resv message to the PHOP signaled earlier in the Path message and which identifies the Ingress Edge. Since the Resv message is directly addressed to the Ingress Edge and does not carry the Router Alert option (as per regular RSVP Resv procedures), the Resv message is hidden from the Interior nodes which handle the E2E Resv message as a regular IP packet.

When receiving the Resv message, the Ingress Edge processes the Resv message as per regular RSVP with the following exceptions:

- 1) if the CL-ECN object is absent from the Resv message, this means that the RSVP Next Hop is not CL-ECN capable and hence proper admission control can not be achieved for that reservation over the ECN cloud. Thus, the Ingress Edge MUST send a ResvErr message towards the receiver with Error Code "Inconsistent Admission Control Behaviour across Ingress and Egress Edge" and an Error Value of "Egress Edge Router not CL-ECN capable". The Ingress Edge MAY also generate an alarm to the network operator.

Note that in the case where the RSVP Next Hop is not CL-ECN capable, this RSVP hop would have (most probably) performed the RSVP-TTL vs IP-TTL check when processing the initial Path message and as a result would have set the Break bit in the Adspec (assuming there is at least one Interior node on the path from the Ingress Edge to the RSVP Next Hop). Thus, the sender would already have been notified in the first place that the QoS could not be guaranteed end-to-end.

- 2) The Ingress Edge MUST carry out the admission control decision (for admission of the reservation over the path from Ingress Edge to Egress Edge through the ECN cloud) taking into account the congestion information provided in the CL-ECN object of the Resv message in accordance with the procedures of [CL-ARCH] and [CL-MARKING] (for example, if the Congestion Level Estimate conveyed in the CL-ECN object exceeds a configured threshold, the Ingress Edge may decide to reject this new reservation). Once the admission control decision is taken by the Ingress Edge, regular RSVP procedures are followed to either proceed with the reservation (and forward the Resv towards the sender) or tear down the reservation (and, in particular, send a ResvErr towards the receiver with Error Code "Admission Control failure").
- 3) In case the Ingress Edge forwards the Resv message upstream, the Ingress Edge MUST remove the CL-ECN object [Editor Note: discuss RSVP Authentication]

When generating a refresh for a Resv message towards the Ingress Edge, the Egress Edge SHOULD NOT include the current value of the Congestion-Level-Estimate in the CL-ECN object, but rather SHOULD include the value which was included in the previous refresh. This is for implementation reasons, to facilitate detection by the Ingress Edge that this message is a mere refresh even if the value of the actual Congestion-Level-Estimate has changed since the previous refresh.

When receiving a PathErr message with the new Error Code of "CL-ECN Probes Required", the Ingress Edge MUST generate CL-ECN probes as described in [CL-ARCH] and [CL-MARKING] towards the Egress Edge which

sent the PathErr Message, and MUST not propagate the PathErr message further upstream.

### [3.3.](#) Removal of E2E reservations

E2E reservations are removed in the usual RSVP way via PathTear, ResvTear, timeout, or as the result of an error condition. This does not directly affect CL-ECN operations.

### [3.4.](#) Overview of Procedures for Preemption of Existing Reservations

As mentioned earlier, [\[CL-ARCH\]](#) describes how the framework uses rate-based pre-emption to maintain the CL service to as many admitted microflows as possible even after localised failure and routing changes in the interior of the edge-to-edge region. The solution has two aspects. First, triggering the Ingress Edge that pre-emption may be needed. This involves a new router marking behaviour for Alert Notification. Secondly, calculating the right amount of traffic to drop. This involves the Egress Edge measuring the amount of CL traffic from a particular ingress gateway. The Ingress edge compares this measurement (which is the amount that the network can actually support) with the rate that it is sending.

This section describes RSVP operations for support of such preemption.

[Section 3.2.1](#) "Alerting the Ingress Edge that pre-emption may be needed" specifies how the Egress Edge can be alerted, explicitly or implicitly, that preemption may be needed and specifies how the Egress Edge measures the rate of traffic at the Egress Edge for the relevant CL-region-aggregate. It also specifies that the Egress Edge needs to transport this information reliably to the Ingress Edge.

Let us assume that a number of reservations are established and transit through a given Ingress Edge  $E_i$  and a given Egress Edge  $E_e$ . Let us now assume that  $E_e$  is alerted that preemption may be needed and that  $E_e$  has measured the egress rate for the CL-region-aggregate from  $E_i$  to  $E_e$ .

Then,  $E_e$  MUST arbitrarily select one of the reservations whose Previous Hop is  $E_i$  and address to  $E_i$  a Resv message for that reservation with a CL-ECN object containing the current Egress Rate for the relevant CL-region-aggregate.

To avoid the risk that this Resv message gets lost and in turn that the Ingress Edge is not made aware in a timely manner that the Congestion-Level-Estimate has changed significantly, the RSVP

reliable messaging procedures specified in [\[RSVP-REFRESH\]](#) SHOULD be used.

Note that, even when reliable messaging is used, there is a very small risk that the information about significant change of Congestion-Level-estimate does not make it to the Ingress Edge. For example, this could happen because there could be a race condition whereby the corresponding reservation may get torn down around the same time where the Resv message with the CL-ECN object is transmitted, resulting in the Ingress Edge ignoring the whole Resv message. However, this appears very unlikely and could also be mitigated by the Egress Edge sending the new Congestion-Level-Estimate on more than one reservation.

[Editor Note: optional use of a Notify message will be investigated. Can this solve the race condition problem mentioned above?]

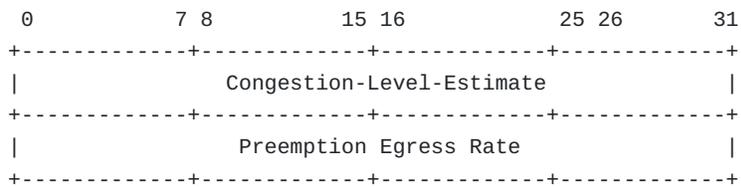
On receipt of the Resv message Ei will detect that this message is not just a refresh because the content of the CL-ECN object has changed and will immediately trigger its admission control logic. This will assess whether some reservations need to be dropped in accordance with the [\[CL-ARCH\]](#) and [\[CL-MARKING\]](#) scheme. In case some do, those will be torn down as per regular RSVP procedures (in particular a ResvErr message is then sent to the receiver).

#### [4.](#) RSVP Object and Error Code Definition

This document defines a new object and two new error codes.

##### [4.1.](#) CL-ECN Object

- o Class = To be allocated by IANA  
C-Type = 1



The CL-ECN Object may only be used in Resv messages.

Let us refer:

- to the Egress Edge which generated the Resv message containing the CL-ECN object as Ee

- to the RSVP Previous HOP (Ingress Edge) for the corresponding reservation as Ei.

#### CL-ECN Congestion-Level-Estimate:

This contains the current value of the Congestion-Level-Estimate (defined in [CL-ARCH] and [CL-MARKING]) computed by Ee for traffic from Ei to Ee

[Editor Note: Encoding details to be added]

#### Preemption Egress Rate:

This contains:

- When Ee is not alerted that preemption is needed for traffic from Ei to Ee, this field is set to 0,
- When Ee is alerted that preemption is needed for traffic from Ei to Ee, the current value of the rate of traffic at the Egress Edge for the relevant CL-region-aggregate (defined in [CL-ARCH] and [CL-MARKING]) computed by Ee for traffic from Ei to Ee

[Editor Note: Encoding details to be added]

### 4.2. "CL-ECN Probes Required" Error Code

The "CL-ECN Probes Required" Error Code may appear only in PathErr messages.

Error Code = To be allocated by IANA

### 4.3. "Inconsistent Admission Control Behaviour across Ingress and Egress Edge" Error Code

The "Inconsistent Admission Control Behaviour across Ingress and Egress Edge" may appear only in ResvErr messages.

[Editor note: should we allow it in PathErr messages too so that notification can also be provided to the sender?]

Error Code for "Inconsistent Admission Control Behaviour across Ingress and Egress Edge"= To be allocated by IANA

Error Value for "Egress Edge Router not CL-ECN capable"= To be allocated by IANA

## 5. Security Considerations

To be added

## 6. IANA Considerations

This document makes the following requests to the IANA:

- allocate a new Object Class (CL-ECN Object)
- allocate a new Error Code ("CL-ECN Probes Required") and manage

the corresponding Error Value range

- allocate a new Error Code ("Inconsistent Admission Control Behaviour across Ingress and Egress Edge") and manage the corresponding Error Value range

## 7. Acknowledgments

We would like to thank Carol Iturralde for her input into this document.

## 8. Normative References

[RSVP] Braden, R., ed., et al., "Resource ReSerVation Protocol (RSVP)- Functional Specification", [RFC 2205](#), September 1997.

[CL-ARCH] Briscoe et al." A Framework for Admission Control over Diffserv using Pre-Congestion Notification", [draft-briscoe-tsvwg-cl-architecture-01.txt](#), October 2005.

[RFC2998] Bernet, Y., Yavatkar, R., Ford, P., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J. and E. Felstaine, "A Framework for Integrated Services Operation Over DiffServ Networks", [RFC 2998](#), November 2000.

[CL-MARKING] work in progress, to be published.

[RSVP-REFRESH] Burger et al, "RSVP Refresh Overhead Reduction Extensions", [RFC2961](#), April 2001

## 9. Informative References

[RFC2211] J. Wroclawski, Specification of the Controlled-Load Network Element Service, September 1997

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "A framework for Differentiated Services", [RFC 2475](#), December 1998.

## 10. Authors Address:

Francois Le Faucheur  
Cisco Systems, Inc.  
Village d'Entreprise Green Side - Batiment T3  
400, Avenue de Roumanille  
06410 Biot Sophia-Antipolis  
France  
Email: flefauch@cisco.com

Anna Charny  
Cisco Systems  
300 Apollo Drive  
Chelmsford, MA 01824  
USA  
EMail: acharny@cisco.com

Bob Briscoe  
BT Research  
B54/77, Sirius House  
Adastral Park  
Martlesham Heath  
Ipswich, Suffolk  
IP5 3RE  
United Kingdom  
Email: bob.briscoe@bt.com

Philip Eardley  
BT Research  
B54/77, Sirius House  
Adastral Park  
Martlesham Heath  
Ipswich, Suffolk  
IP5 3RE  
United Kingdom  
Email: philip.eardley@bt.com

Kwok Ho Chan  
Nortel Networks  
600 Technology Park Drive  
Billerica, MA 01821  
USA  
Email: khchan@nortel.com

Jozef Z. Babiarz  
Nortel Networks  
3500 Carling Avenue  
Ottawa, Ont K2H 8E9  
Canada  
Email: babiarz@nortel.com

## 11. IPR Statements

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](mailto:ietf-ipr@ietf.org).

## 12. Disclaimer of Validity

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.

## 13. Copyright Notice

Copyright (C) The Internet Society (2005). 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.

## Appendix A - Example RSVP Signaling Flow for Admission Control

To be added. Shows RSVP message flow in case of admission control of new reservations.

Appendix B - Example Signaling Flow for Preemption

To be added. Shows RSVP message flow in case of preemption of existing reservations.