

Networking Working Group
Internet Draft

Reshad Rahman
Anca Zamfir
Junaid Ismail
Cisco Systems

Document: [draft-rahman-ccamp-rsvp-restart-extensions.txt](#)

Expires: August 2004

February 2004

RSVP Graceful Restart Extensions

[draft-rahman-ccamp-rsvp-restart-extensions-00.txt](#)

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 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 describes the extensions needed by certain features for the purpose of RSVP Graceful Restart (defined in [RFC3473]). One of these extensions refers to the ability of a node to recover the ERO in the case it has performed an ERO expansion before control plane restart. Also a small

[draft-rahman-ccamp-rsvp-restart-extensions-00.txt](#)

February 200

modification is proposed in the basic procedure to support simultaneous multiple node restarts in a network. Specifically, a node should use a non-zero Recovery Time while in the recovery phase. This allows a node to determine at restart time if any of its neighbors has previously restarted and it is currently in the recovery phase.

Conventions used in this document

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 [RFC 2119](#) [[RFC2119](#)].

Routing ID Summary

(This section to be removed before publication.)

SUMMARY

This document specifies extensions and mechanisms to RSVP Graceful Restart to provide support for ERO Recovery and multiple node restart.

WHERE DOES IT FIT IN THE PICTURE OF THE ROUTING WORK?

This work fits in the context of [[RFC3473](#)].

WHY IS IT TARGETED AT THIS WG?

This draft is targeted at this WG, because it specifies extensions to RSVP graceful restart which is defined in [[RFC3473](#)].

RELATED REFERENCES

Please refer to the reference section.

Table of Contents

1. Terminology

GR – Graceful Restart procedure for RSVP as specified in [\[RFC3473\]](#).

2.

RSVP Graceful Restart

The procedure for RSVP Graceful Restart (GR) is described in [\[RFC3473\]](#). The purpose of this procedure is to allow a node that has experienced a failure in the control plane but that has preserved its forwarding plane to recreate its states based on replayed RSVP messages received from its neighbors and also based on information retrieved from the forwarding plane. Typically, the data that can be obtained from the forwarding plane for each LSP endpoint is (port, label). At mid, control plane obtains also the cross-connect information: (ingress-port, ingress-label) X (egress-port, egress-label)

While most of the objects can be recovered based on these two sources of information, the route information that is contained in the ERO object cannot be recovered if the restarting node has modified the ERO content prior to restart. [Section 3](#) proposes a solution for this problem.

The GR procedure described in [\[RFC3473\]](#) handles some cases of multiple node restart. In the example below, assume that one LSP was signaled by R1 to span L1, R2, L2, R3.

```
      L1      L2
R1 ----- R2 ----- R3
```

If R2 restarts and then R3 restarts after the Hellos have come up, then as described in [\[RFC3473\]](#), R2 is able to detect that R3 has restarted and if R2 is in the recovery mode when R3 has restarted, when R2 receives a recovery message from R1 it sends a recovery message to R3. For the LSP described before, R1 sends the Path message including a Recovery Label and R2 also includes a Recovery Label in the Path message sent to R3.

If R2 restarts and then R3 restarts before the Hellos have come up, then there is no specified way for R2 to detect that R3 has restarted and is currently in Recovery Mode. Therefore, when R2 receives the Path message with the Recovery Label from R1, after processing it, R2 sends the corresponding outgoing Path message to R3 with Suggested Label instead of the Recovery Label. This is incorrect since this message must include the Recovery Label in order to help R3 recover its state.

A solution for this problem is described in [Section 4](#).

3.

ERO Recovery

If a node experiences a control plane failure and restarts, the existing GR procedures do not ensure that ERO expansion before and after the failure yield the same results. A change in ERO expansion should be avoided as it may lead to undesirable results.

This section describes how such a change can be prevented. To support this solution, a new RSVP object, called Recovery ERO is introduced.

3.1

Recovery ERO Object

The Recovery ERO object is used during nodal fault recovery process. The format of Recovery ERO object is identical to that of the ERO object described in [\[RFC3209\]](#). A Recovery ERO object uses Class-Number ?? (of form 10bbbbbb) and the same C-Type as the one of the ERO object it is trying to recover. Only C-Type = 1 is currently supported.

3.2

Procedures at the Restarting Node and its Neighbors

When a node experiences control plane restart and receives a

Path message with Recovery Label from the upstream node, it searches for a matching forwarding state as per [RFC3473]. If no matching state is found and if ERO expansion is required, then the node considers the Path message as a new LSP. It processes the incoming Path message and performs ERO expansion as specified in [RFC3473], [RFC3209]. If the forwarding state is found and if ERO expansion is required, then the node processes the incoming Path message as specified in [RFC3473], [RFC3209] with following modifications:

1. It performs partial ERO expansion at this point to include:
 - the strict next hop that is contained in the forwarding state.
 - the loose hop as in the ERO of the received Path message.
2. It includes the result of the previous step in the Recovery ERO object to be sent as part of the outgoing Path message.
3. The restarting node sends the outgoing Path message out.

When the Path message is received by the neighbor downstream of the restarting node, the following processing occurs:

4. If this message has associated incoming Path and forwarding states, the neighbor node retrieves the ERO object as it was previously created by the restarting node and formats a new Recovery ERO object with this content to be sent upstream.

5. If this message has neither an associated incoming Path state nor a forwarding state, then this should be treated as a normal setup.

6. If this message has no associated Path state but forwarding state is present, then this node is restarting as well and the procedure of the restarting node applies. Once the outgoing Path state is recovered, this node retrieves the outgoing ERO and creates the Recovery ERO object by prepending one or more strict elements as identified by forwarding entry associated with this LSP on this abstract node.

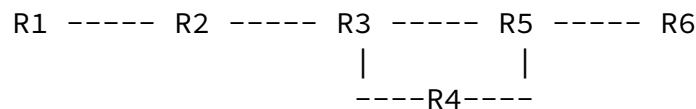
If a new upstream Recovery ERO object is available after executing the steps 4, 5 and 6, then the neighbor node includes the upstream ERO content in a Recovery ERO object to be sent upstream in the Resv message.

When the restarting node receives the Resv message (after step 3), it removes the Recovery ERO object before creating the Reservation State and uses its content to update the ERO in the associated Path State.

In the case where restarting node determines that the downstream node has not been able to include the expanded ERO (e.g. downstream node is also a restarting node and forwarding has not been preserved), the restarting node performs the expansion as described in [[RFC3473](#)], [[RFC3209](#)]. In this case the recovery of the LSP is not guaranteed.

Below is an example that covers the steps above:

Assume a simple topology as follows:



1. R1 sends a Path message to R2 with an ERO containing [R2, R6(loose)]. R2 performs ERO expansion [R3, R4, R5, R6] and forwards the Path message to R3. R3 stores this ERO and LSP gets established using normal LSP setup procedures.
2. The control plane on R2 restarts.
3. R2 receives a Path message with Recovery Label and ERO = [R2, R6(loose)]. R2 finds the forwarding state and creates the incoming Path state with ERO = [R2, R6].

4. Based on the forwarding state, R2 determines that the next hop is R3. R2 creates an outgoing Path State with a Recovery ERO = [R2, R3, R6] and forwards the Path message to R3.
5. R3 finds the associated incoming Path State with ERO = [R3, R4, R5, R6] and creates a Recovery ERO with this content. R3 sends the Resv message upstream including the Recovery ERO object that contains: [R3, R4, R5, R6]
6. R2 receives the Resv message, removes the Recovery ERO

object and creates the Reservation state.

7. R2 uses the Recovery ERO from the Resv message to create the ERO of the outgoing Path State as [R3, R4, R5, R6] and removes the Recovery ERO.

4.

Handling Multiple Restarts

If a node (R2 below) experiences a control plane failure and if this node implements Graceful Restart procedure described in [\[RFC3473\]](#), then it can correctly recover all RSVP states it had prior to restart.

During recovery, new LSPs may be established as long as they do not collide with the LSPs that are in progress of being recovered. If a Path message that includes a Suggested Label is received and if the restarting node checks its forwarding and determines that a previous LSP was using the (ingress-port, ingress-label), the new request may get rejected. It may also get accepted with an upstream label different than the one suggested in the Suggested Label.

R0 ----- R1 ----- R2 ----- R3

With the current specification, if a node R1 upstream of the restarting node R2 is also in the recovery phase, then the only case where the LSP is recovered is when R1 has restarted prior to R2. In this case R1 determines based on the Hello session that R2 has restarted and therefore it sends Path Messages with Recovery Label for the LSPs that need to be recovered.

In the case R1 restarted after R2, R1 cannot determine based on the Hello Instance that R2 is a restarting node. When R1 receives recovery Path messages from its upstream node, it sends Path messages with Suggested Label to R2 as indicated in

[\[RFC3473\]](#). R2 does not use the Suggested Label in this case. This is because its forwarding indicates that a Recovery Path

message may be received from R1. In fact, it may setup a new LSP using a different label.

In the case of GMPLS, messages from different upstream neighbors may be received on the same interface which may be different from the interface hosting the LSP to be recovered.

This draft proposes the use of Recovery Time value received by a restarting node as an indication that its downstream neighbor is in recovery mode.

A node that complies with this specification sends non-zero Recovery Time while in recovery mode and MUST set it to 0 once its recovery has completed.

The alternative solution is to require restarting nodes that receive a Path with Recovery Label to forward, after processing, a Path with Recovery Label and not Suggested Label.

To allow for recovery to complete, a restarting node may wish to adjust its advertised Recovery Time when an upstream node restarts, in an attempt to prevent the downstream nodes to expire states that may be recovered later than expected.

5.

Forward Compatibility Note

A node that does not support the Recovery ERO object and the procedure described in this draft, will ignore the Recovery ERO object and responds with a Resv. A restarting node may choose to continue the recovery by performing a new ERO expansion. In the case where the new ERO matches the ERO before restart the LSP is recovered. Otherwise, depending on the downstream node implementations, the LSP may be torn down.

The extensions specified in this draft do not affect the processing of the Restart Cap object at nodes that do not support them.

A node that does not comply with this specification and has no proprietary way to detect at restart time the downstream neighbors that have previously restarted and that are in the recovery mode, may ignore the Recovery Time in the Restart_Cap object and may forward only Path messages with Suggested Label.

A node that does comply with this specification and that receives a Restart_Cap object with a non-zero Recovery Time from a downstream node that does not comply with this specification, forwards Path messages with Recovery Label included for all recovered LSPs while in the recovery period. If in fact the downstream node is not in Recovery mode and receives a Path message with a Recovery Label, it should generate a Resv message and normal state processing continues.

6.

Security Considerations

This document does not introduce new security issues. The security considerations pertaining to the original RSVP protocol [[RFC2205](#)] remain relevant.

References

- [[RFC2205](#)] "Resource ReSerVation Protocol (RSVP) - Version 1, Functional Specification", [RFC 2205](#), Braden, et al, September 1997.
- [[RFC3209](#)] "Extensions to RSVP for LSP Tunnels", D. Awduche, et al, [RFC 3209](#), December 2001.
- [[RFC3471](#)] "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", [RFC 3471](#), L. Berger, et al, January 2003.
- [[RFC3473](#)] "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3471](#), L. Berger, et al, January 2003.
- [[RFC3477](#)] "Signaling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE) ", [RFC 3477](#), K. Kompella, Y. Rekhter, January 2003.
- [[RFC2119](#)] "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), S. Bradner, March 1997.

Author's Addresses

Reshad Rahman
Cisco Systems Inc.
2000 Innovation Dr.,
Kanata, Ontario, K2K 3E8

Canada.
Phone: (613)-254-3519
Email: rrahman@cisco.com

Rahman, R., Zamfir, A., Israr, J.

[Page 8

[draft-rahman-ccamp-rsvp-restart-extensions-00.txt](#)

February 200

Anca Zamfir
Cisco Systems Inc.
2000 Innovation Dr.,
Kanata, Ontario, K2K 3E8
Canada.
Phone: (613)-254-3484
Email: ancaz@cisco.com

Junaid Israr
Cisco Systems Inc.
2000 Innovation Dr.,
Kanata, Ontario, K2K 3E8
Canada.
Phone: (613)-254-3693
Email: jisrar@cisco.com

