Network Working Group
Internet Draft
Category: Standards Track
Expires: April 2004

Adrian Farrel (editor)
Old Dog Consulting

Arun Satyanarayana
Movaz Networks, Inc.

Atsushi Iwata
Norihito Fujita
NEC Corporation

Gerald R. Ash
AT&T

Simon Marshall-Unitt
Data Connection Ltd.

October 2003

Crankback Signaling Extensions for MPLS Signaling
<draft-iwata-mpls-crankback-07.txt>

Status of this Memo

Abstract

In a distributed, constraint-based routing environment, the information used to compute a path may be out of date. This means that Multiprotocol Label Switching (MPLS) label switched path (LSP) setup requests may be blocked by links or nodes without sufficient resources. Crankback is a scheme whereby setup failure information is returned from the point of failure to allow new setup attempts to be made avoiding the blocked resources. Crankback can also be applied to LSP restoration to indicate the location of the failed link or node.

This document specifies crankback signaling extensions for use in MPLS signaling using RSVP-TE as defined in "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, so that the LSP setup request can be retried on an alternate path that detours around blocked links or nodes. This offers significant improvements in the successful setup and recovery ratios for LSPs, especially in situations where a large number of setup requests are triggered at the same time.

Table of Contents

Section A : Problem Statement

1. Summary for Sub-IP Area

1.1. Summary

This document describes requirements, procedures and protocol extensions for Crankback Routing in MPLS and GMPLS networks. These extensions address some of the requirements laid out by the ITU-T for the Automatically Switched Optical Network (ASON). This is recognized in [ASON-REQ].

1.2. Related documents

See the References Sections.

1.3. Where does it fit in the Picture of the Sub-IP Work

This work is applicable to MPLS and GMPLS signaling protocols.

1.4. Why is it Targeted at this WG

MPLS is a product of the MPLS WG, GMPLS is worked on by the CCAMP WG. This document provides common extensions for use in MPLS and GMPLS and so is appropriate for consideration by the CCAMP WG.

The CCAMP charter now contains the work item:

- Define signaling and routing mechanisms to make possible the creation of paths that span multiple IGP areas, multiple ASes, and multiple providers, including techniques for crankback.

1.5. Justification

Crankback Signaling is a requirement in large and multi-area networks, in networks with rapidly changing topologies or resource usage, or in networks where setup latency may be high.

The requirement for Crankback Routing in the Automatically Switched Optical Network (ASON) has been identified by the ITU-T [G8080] and recognized by the IETF in [ASON-REQ].

2. Introduction and Framework

2.1. Background

RSVP-TE (RSVP Extensions for LSP Tunnel) [RFC3209] can be used for establishing explicitly routed LSPs in an MPLS network. Using RSVP-TE, resources can also be reserved along a path to guarantee or control QoS for traffic carried on the LSP. To designate an explicit path that satisfies QoS constraints, it is necessary to discern the resources available to each link or node in the network. For the collection of such resource information, routing protocols, such as OSPF and IS-IS , can be extended to distribute additional state information [RFC2702].

Explicit paths can be computed based on the distributed information at the LSR initiating a LSP and signaled as Explicit Routes during LSP establishment. Explicit Routes may contain 'loose hops' and 'abstract nodes' that convey

routing through any of a collection of nodes. This mechanism may be used to devolve parts of the path computation to intermediate nodes such as area border LSRs.

In a distributed routing environment, however, the resource information used to compute a constraint-based path may be out of date. This means that a setup request may be blocked, for example, because a link or node along the selected path has insufficient resources.

In RSVP-TE, a blocked LSP setup may result in a PathErr message sent to the initiator or a ResvErr sent to the terminator (egress LSR). These messages may result in the LSP setup being abandoned. In Generalized MPLS [RC3473] the Notify message may additionally be used to expedite notification of LSP failures to ingress and egress LSRs, or to a specific "repair point".

These existing mechanisms provide a certain amount of information about the path of the failed LSP.

2.2. Repair and Restoration

If the ingress LSR or intermediate area border LSR knows the location of the blocked link or node, the LSR can designate an alternate path and then reissue the setup request. Determination of the identity of the blocked link or node can be achieved by the mechanism known as crankback routing [PNNI, ASH1]. In RSVP-TE, crankback signaling requires notifying an upstream LSR of the location of the blocked link or node. In some cases this requires more information than is currently available in the signaling protocols.

On the other hand, various restoration schemes for link or node failures have been proposed in [RFC3469] and others including fast restoration. These schemes rely on the existence of a backup LSP to protect the primary, but

if both the primary and backup paths fail it is necessary to reestablish the LSP on an end-to-end basis avoiding the known failures. Similarly, fast restoration by establishing a restoration path on demand after failure requires computation of a new LSP that avoids the known failures. End-to-end restoration for alternate routing requires the location of the failed link or node. Crankback routing schemes could also be used to notify

upstream LSRs of the location of the failure.

Furthermore, in situations where many link or node failures occur at the same time, the difference between the distributed routing information and the real-time network state becomes much greater than in normal LSP setups. LSP restoration might, therefore, be performed with inaccurate information, which is likely to cause setup blocking. Crankback routing could improve failure recovery in these situations.

Generalized MPLS [RFC3471] extends MPLS into networks that manage Layer2, TDM and lambda resources. In a network without wavelength converters, setup requests are likely to be blocked more often than in a conventional MPLS environment because the same wavelength must be allocated at each Optical Cross-Connect on an end-to-end explicit path. Furthermore, end-to-end restoration is the only way to recover LSP failures. This implies that crankback routing would also be useful in a GMPLS network, in particular in dynamic LSP re-routing cases (no backup LSP pre-establishment).

3. Discussion: Explicit Versus Implicit Re-routing Indications

There have been problems in service provider networks when "inferring" from indirect information that re-routing is allowed. This document proposes the use of an explicit re-routing indication that explicitly authorizes re-routing.

Various existing protocol options and exchanges including the error values of PathErr message [RFC2205, RFC3209] and the Notify message [RFC3473] allow an implementation to infer a situation where re-routing can be done. This allows for recovery from network errors or resource contention.

However, such inference of recovery signaling is not always desirable since it may be doomed to failure. Experience of using release messages in TDM-based networks for analogous purposes provides some guidance. One can use the receipt of a release message with a cause value (CV) indicating "link congestion" to trigger a re-routing attempt at the originating node. However, this sometimes leads to problems.

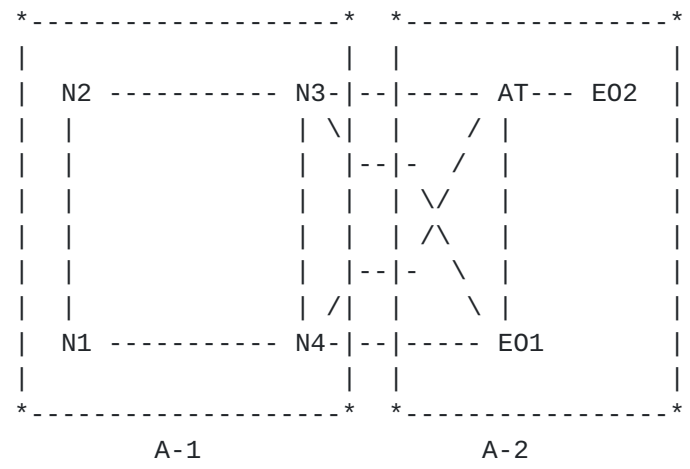


Figure 1. Example of network topology

Figure 1 illustrates four examples based on service-provider experiences with respect to crankback (i.e., explicit indication) versus implicit indication through a release with CV. In this example, N1, N2,N3, and N4 are located in one area (A-1), and AT, EO1, and EO2 are in another area (A-2).

Note that two distinct areas are used in this example to expose the issues clearly. In fact, the issues are not limited to multi-area networks, but arise whenever path computation is distributed throughout the network. For example where loose routes, AS routes or path computation domains are used.

1. A connection request from node N1 to EO1 may route to N4 and then find "all circuits busy". N4 returns a release message to N1 with CV34 indicating all circuits busy. Normally, a node such as N1 is programmed to block a connection request when receiving CV34, although there is good reason to try to alternate route the connection request via N2 and N3.

   Some service providers have implemented a technique called route advance (RA), where if a node that is RA capable receives a release message with CV34, it will use this as an implicit re-route indication and try to find an alternate route for the connection request if possible. In this example, alternate route N1-N2-N3-EO1 can be tried and may well succeed.

2. Suppose a connection request goes from N2 to N3 to AT trying to reach EO2 and is blocked at link AT-EO2. Node AT returns a CV34 and with RA, N2 may try to re-route N2-N1-N4-

AT-EO2, but of course this fails again. The problem is that N2 does not realize where this blocking occurred based on the CV34, and in this case there is no point in further alternate routing.

3. However, in another case of a connection request from N2 to E02, suppose that link N3-AT is blocked. In this case N3 should return crankback information (and not CV34) so that N2 can alternate route to N1-N4-AT-EO2, which may well be successful.

4. In a final example, for a connection request from EO1 to N2, EO1 first tries to route the connection request directly to N3. However, node N3 may reject the connection request even if there is bandwidth available on link N3-EO1 (perhaps for priority routing considerations, e.g., reserving bandwidth for high priority connection requests). However, when N3 returns CV34 in the release message, EO1 blocks the connection request (a normal response to CV34 especially if E01-N4 is already known blocked) rather than trying to alternate route through AT-N3-N2, which might be successful. If N3 returns crankback information, EO1 could respond by trying the alternate route.

It is certainly the case that with topology exchange, such as OSPF, the ingress LSR could infer the re-routing condition. However, convergence of routing information is typically slower than the expected LSP setup times. One of the reasons for crankback is to avoid the overhead of available-link-bandwidth flooding, and to more efficiently use local state information to direct alternate routing at the ingress-LSR.

[ASH1] shows how event-dependent-routing can just use crankback, and not available-link-bandwidth flooding, to decide on the re-route path in the network through "learning models". Reducing this flooding reduces overhead and can lead to the ability to support much larger AS sizes.

Therefore, the alternate routing should be indicated based on an explicit indication (as in examples 3 and 4), and it is best to know the following information separately:

a) where blockage/congestion occurred (as in examples 1-2),

and

b) whether alternate routing "should" be attempted even if there is no "blockage" (as in example 4).

4. Required Operation

Section 2 identifies some of the circumstances under which crankback may be useful. Crankback routing is performed as described in the following procedures, when an LSP setup request is blocked along the path or when an existing LSP fails.

4.1. Resource Failure or Unavailability

When an LSP setup request is blocked due to unavailable resources, an error message response with the location identifier of the blockage should be returned to the LSR initiating the LSP setup (ingress LSR), the area border LSR, the AS border LSR, or to some other repair point.

This error message carries an error specification according to [RFC3209] - this indicates the cause of the error and the node/link on which the error occurred. Crankback operation may require further information as detailed in section 6.

4.2. Computation of an Alternate Path

In a flat network without partitioning, when the ingress LSR receives the error message it computes an alternate path around the blocked link or node to satisfy QoS constraints using link state information about the area. If an alternate path is found, a new LSP setup request is sent over this path.

On the other hand, in a network partitioned into areas such as with hierarchical OSPF, an area border LSR may intercept and terminate the error response, and perform alternate (re-)routing within the downstream area.

In a third scenario, any node within an area may act as a

repair point. In this case, the LSR behaves much as an area border LSR as described above. It can intercept and terminate the error response, and perform alternate routing. This may be particularly useful where domains of computation are applied within the network, however if all nodes in the network perform re-routing it is possible to spend excessive network and CPU resources on re-routing attempts that would be better made only at designated re-routing nodes. This scenario is somewhat like 'MPLS fast re-route' [FASTRR], in which any node in the MPLS domain can establish 'local repair' LSPs after failure notification.

4.2.1 Information Required for Re-routing

In order to correctly compute a route that avoids the blocking problem, a repair point LSR must gather as much crankback information as possible. Ideally, the repair node will be given the node, link and reason for the failure.

However, this information may not be enough to help with re-computation. Consider for instance an explicit route that contains a non-explicit abstract node or a loose hop. In this case, the failed node and link is not necessarily enough to tell the repair point which hop in the explicit route has failed. The crankback information needs to provide the context into the explicit route.

4.2.2 Signaling a New Route

If the crankback information can be used to compute a new route avoiding the blocking problem, the route can be signaled as an Explicit Route.

However, it may be that the repair point does not have sufficient topology information to compute an Explicit Route that is guaranteed to avoid the failed link or node. In this case, Route Exclusions [EXCLUDE] may be particularly helpful. To achieve this, [EXCLUDE] allows the crankback information to be presented as route exclusions to force avoidance of the failed node, link or resource.

4.3. Persistence of Error Information

The repair point LSR that computes the alternate path should store the location identifiers of the blockages indicated in the error message until the LSP is

successfully established or until the LSR abandons re-routing attempts. Since crankback routing may happen more than once while establishing a specific LSP, a history table of all experienced blockages for this LSP SHOULD be maintained (at least until the routing protocol updates the state of this information) to perform an accurate path computation to detour all blockages.

If a second error response is received by a repair point (while it is performing crankback re-routing) it should update the history table that lists all experienced blockages, and use the entire gathered information when making a further re-routing attempt.

4.4. Handling Re-route Failure

Multiple blockages (for the same LSP) may occur, and successive setup retry attempts may fail. Retaining error information from previous attempts ensures that there is no thrashing of setup attempts, and knowledge of the blockages increases with each attempt.

It may be that after several retries, a given repair point is unable to compute a path to the destination (that is, the egress of the LSP) that avoids all of the blockages. In this case, it must pass the error indication upstream. It is most useful to the upstream nodes (and in particular the ingress LSR) that may, themselves, attempt new routes for the LSP setup if the error indication in this case identifies all of the downstream blockages and also the node that has been unable to compute an alternate path.

4.5. Limiting Re-routing Attempts

It is important to prevent an endless repetition of LSP setup attempts using crankback routing information after error conditions are signaled, or during periods of high congestion. It may also be useful to reduce the number of retries, since failed retries will increase setup latency and degrade performance.

The maximum number of crankback re-routing attempts allowed may be limited in a variety of ways. The number may be limited by LSP, by node, by area or by AS. Control of the limit may be applied as a configuration item per LSP, per node, per area or per AS.

When the number of retries at a particular node, area or AS is exceeded, the LSR handling the current failure reports the failure upstream to the next node, area or AS where further re-routing attempts may be attempted. It is important that the crankback information provided indicates that routing back through this node, area or AS will not succeed - this situation is similar to that in section 4.4. Note that in some circumstances, such a report will also mean that no further re-routing attempts can possibly succeed - for example, when the egress node is within the failed area.

When the maximum number of retries for a specific LSP has been exceeded, the LSR handling the current failure should send an error message upstream indicating "Maximum number of re-routings exceeded". This error will be passed back to the ingress LSR with no further re-routing attempts. The ingress LSR may choose to retry the LSP setup according to local policy and might choose to re-use its original path or seek to compute a path that avoids the blocked resources. In the latter case, it may be useful to indicate the blocked resource in this error message.

5. Existing Protocol Support for Crankback Re-routing

Crankback re-routing is appropriate for use with RSVP-TE.

1) Path establishment may fail because of an inability to route, perhaps because links are down. In this case a PathErr message is returned to the initiator.

2) Path establishment may fail because resources are unavailable. This is particularly relevant in GMPLS where explicit label control may be in use. Again, a PathErr message is returned to the initiator.

3) Resource reservation may fail in the upstream direction, as the Resv is processed, and resources are reserved. If resources are not available on the required link or at a specific node, a ResvErr message is returned to the egress node indicating "Admission Control failure" [RFC2205]. The

egress is allowed to change the FLOWSPEC and try again, but in the event that this is not practical or not supported (particularly in the GMPLS context), the egress LSR may choose to take any one of the following actions.

- Ignore the situation and allow recovery to happen through Path refresh message and refresh timeout [RFC2205].
- Send a PathErr message towards the initiator indicating "Admission Control failure".
- Send a ResvTear message towards the initiator to abort the LSP setup.

Note that in multi-area networks, the ResvErr might be intercepted and acted on at an area border router.

4) It is also possible to make resource reservations on the forward path as the Path message is processed. This choice is compatible with LSP setup in GMPLS networks [RFC3471]. In this case if resources are not available, a PathErr message is returned to initiator indicating "Admission Control failure".

Crankback information would be useful to an upstream node (such as the ingress) if it is supplied on a PathErr or a Notify message that is sent upstream.

5.1. RSVP-TE [RFC 3209]

In RSVP-TE a failed LSP setup attempt results in a PathErr message returned upstream. The PathErr message carries an ERROR_SPEC object, which indicates the node or interface reporting the error and the reason for the failure.

Crankback re-routing can be performed explicitly avoiding the node or interface reported.

5.2. GMPLS-RSVP-TE [RFC 3473]

GMPLS extends the error reporting described above by allowing LSRs to report the interface that is in error in addition to the identity of the node reporting the error. This further enhances the ability of a re-computing node to route around the error.

GMPLS introduces a targeted Notify message that may be used to report LSP failures direct to a selected node. This message carries the same error reporting facilities as described above. The Notify message may be used to expedite the propagation of error notifications, but in a network that offers crankback routing at multiple nodes

there would need to be some agreement between LSRs as to whether PathErr or Notify provides the stimulus for crankback operation. Otherwise, multiple nodes might attempt to repair the LSP at the same time, in particular because 1) these messages can flow through different paths before reaching the ingress LSR and 2) the destination of the Notify message might not be the ingress LSR.

Section B : Solution

6. Control of Crankback Operation

6.1. Requesting Crankback and Controlling In-Network Re-routing

When a request is made to set up an LSP tunnel, the ingress LSR should specify whether it wants crankback information to be collected in the event of a failure and whether it requests re-routing attempts by any or specific intermediate nodes. For this purpose, a Re-routing Flag field is added to the protocol setup request messages. The corresponding values are mutually exclusive.

No Re-routing	Intermediate nodes SHOULD NOT attempt re-routing after failure. Nodes detecting failures MUST report an error and MAY supply crankback information. This is the default and backwards compatible option.
End-to-end Re-routing	Intermediate nodes SHOULD NOT attempt re-routing after failure. Nodes detecting failures MUST report an error and SHOULD supply crankback information.
Boundary Re-routing	Intermediate nodes MAY attempt re-routing after failure only if they are Area Border Routers or AS Border Routers. The boundary ABR/ASBR can either decide to forward the Path Error message upstream to the Head-end LSR or try to select another egress boundary LSR. Other nodes SHOULD NOT attempt re-routing. Nodes detecting failures MUST report an error and SHOULD supply crankback information.
Segment-based Re-routing	All intermediate nodes MAY attempt re-routing after failure. Nodes detecting failures MUST report an error and SHOULD

supply full crankback information.

6.2. Action on Detecting a Failure

A node that detects the failure to setup an LSP or the failure of an established LSP SHOULD act according to the Re-routing Flag passed on the LSP setup request.

If Segment-based Re-routing is allowed or if Boundary Re-routing is allowed and the detecting node is an ABR or ASBR, the detecting node MAY immediately attempt to re-route.

If End-to-end Re-routing is indicated, or if Segment-based or Boundary Re-routing is allowed and the detecting node chooses not to make re-routing attempts (or has exhausted all possible re-routing attempts), the detecting node returns a protocol error indication and SHOULD include full crankback information.

6.3. Limiting Re-routing Attempts

Each repair point should apply a locally configurable limit to the number of attempts it makes to re-route an LSP. This helps to prevent excessive network usage in the event of significant faults and allows back-off to other repair points which may have a better chance of routing around the problem.

6.3.1 New Status Codes for Re-routing

An error code/value of "Routing Problem"/"Re-routing limit exceeded" (24/TBD) is used to identify that a node has abandoned crankback re-routing because it has reached a threshold for retry attempts.

A node receiving an error response with this status code MAY also attempt crankback re-routing, but it is RECOMMENDED that such attempts be limited to the ingress LSR.

6.4. Protocol Control of Re-routing Behavior

The Session Attributes Object in RSVP-TE is used on Path messages to indicate the capabilities and attributes of the session. This object contains an 8-bit flag field which is used to signal individual Boolean capabilities or attributes. The Re-Routing Flag described in section 5.1 would fit naturally into this field, but there is a scarcity of bits, so use is made of the new LSP_ATTRIBUTES object defined in [LSP-ATTRIB]. Three bits are defined for inclusion in the LSP

Attributes TLV as follows. The values below are suggested and actual values are TBD by IETF consensus.

0x01 End-to-end re-routing desired
This flag indicates the end-to-end re-routing behavior for an LSP under establishment. This MAY also be used for specifying the behavior of end-to-end LSP restoration for established LSPs.

0x02 Boundary re-routing desired.
This flag indicates the boundary re-routing behavior for an LSP under establishment. This MAY also be used for specifying the segment-based (hierarchical) LSP restoration for established LSPs. The boundary ABR/ASBR can either decide to forward the PathErr message upstream to the Head-end LSR or try to select another egress boundary LSR.

0x04 Segment-based re-routing desired.
This flag indicates the segment-based re-routing behavior for an LSP under establishment. This MAY also be used for specifying the segment-based LSP restoration for established LSPs.

7. Reporting Crankback Information

7.1. Required Information

As described above, full crankback information should indicate the node, link and other resources, which have been attempted but have failed because of allocation issues or network failure.

The default crankback information SHOULD include the interface and the node address.

7.2. Protocol Extensions

[RFC3473] defines an IF_ID ERROR_SPEC Object that can be used on PathErr, ResvErr and Notify messages to convey the information carried in the Error Spec Object defined in [RFC 3209]. Additionally, it has scope for carrying TLVs that help identify the identity of the link associated with the error.

The TLVs for use with this object are defined in [RFC3471], and are as follows. They are used to identify links in the IF_ID PHOP Object and in the IF_ID ERROR_SPEC Object to identify the failed resource which is usually the downstream resource from the reporting node.

Type	Length	Format	Description	
1	8	IPv4 Addr.	IPv4	(Interface address)
2	20	IPv6 Addr.	IPv6	(Interface address)
3	12	Compound	IF_INDEX	(Interface index)
4	12	Compound	COMPONENT_IF_DOWNSTREAM	(Component interface)
5	12	Compound	COMPONENT_IF_UPSTREAM	(Component interface)

Two new TLVs are defined for use in the IF_ID PHOP Object and in the IF_ID Error Spec Object. Note that the Type values shown here are only suggested values - final values are TBD and to be determined by IETF consensus.

Type	Length	Format	Description	
6	16	See below	UNUM_COMPONENT_IF_DOWN	(Component interface)
7	16	See below	UNUM_COMPONENT_IF_UP	(Component interface)

In order to facilitate reporting of crankback information, the following additional TLVs are defined. Note that the Type values shown here are only suggested values - final values are TBD and to be determined by IETF consensus.

Type	Length	Format	Description	
8	var	See below	DOWNSTREAM_LABEL	(GMPLS label)
9	var	See below	UPSTREAM_LABEL	(GMPLS label)
10	8	See below	NODE_ID	(Router Id)
11	x	See below	OSPF_AREA	(Area Id)
12	x	See below	ISIS_AREA	(Area Id)
13	8	See below	AUTONOMOUS_SYSTEM	(Autonomous system)
14	var	See below	ERO_CONTEXT	(ERO subobject)
15	var	See below	ERO_NEXT_CONTEXT	(ERO subobjects)

16	8	IPv4 Addr.	PREVIOUS_HOP_IPv4	(Node address)
17	20	IPv6 Addr.	PREVIOUS_HOP_IPv6	(Node address)
18	8	IPv4 Addr.	INCOMING_IPv4	(Interface address)
19	20	IPv6 Addr.	INCOMING_IPv6	(Interface address)
20	12	Compound	INCOMING_IF_INDEX	(Interface index)
21	12	Compound	INCOMING_COMP_IF_DOWN	(Component interface)
22	12	Compound	INCOMING_COMP_IF_UP	(Component interface)
23	16	See below	INCOMING_UNUM_COMP_DOWN	(Component interface)
24	16	See below	INCOMING_UNUM_COMP_UP	(Component interface)
25	var	See below	INCOMING_DOWN_LABEL	(GMPLS label)
26	var	See below	INCOMING_UP_LABEL	(GMPLS label)
27	8	See below	REPORTING_NODE_ID	(Router Id)
28	x	See below	REPORTING_OSPF_AREA	(Area Id)
29	x	See below	REPORTING_ISIS_AREA	(Area Id)
30	8	See below	REPORTING_AS	(Autonomous system)
31	var	See below	PROPOSED_ERO	(ERO subobjects)
32	var	See below	NODE_EXCLUSIONS	(List of nodes)
33	var	See below	LINK_EXCLUSIONS	(List of interfaces)

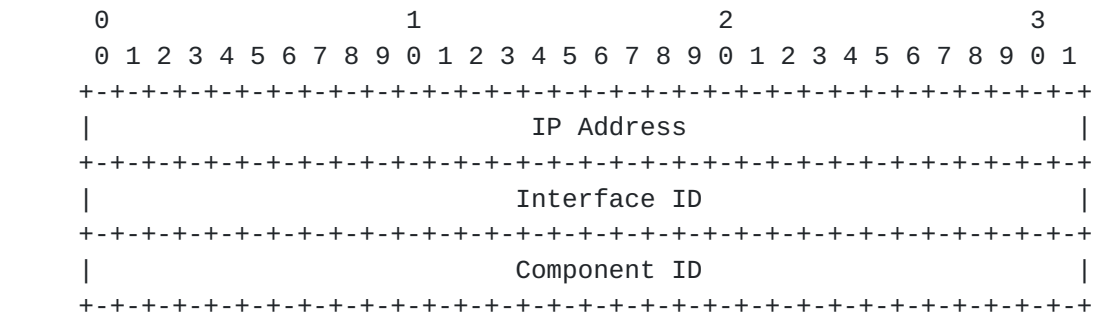
For types 1, 2, 3, 4 and 5, the format of the Value field is already defined in [RFC3471].

For types 16 and 18, they format of the Value field is the same as for type 1.

For types 17 and 19, the format of the Value field is the same as for type 2.

For types 20, 21 and 22, the formats of the Value fields are the same as for types 3, 4 and 5 respectively.

For types 6, 7, 23 and 24 the Value field has the format:



IP Address: 32 bits

The IP address field may carry either an IP address associated with the router, where associated address is the value carried in a router address TLV of routing.

Interface ID: 32 bits

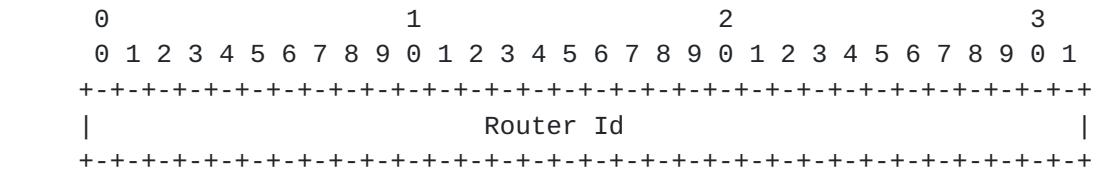
The Interface ID identifier of the unnumbered link.

Component ID: 32 bits

A bundled component link. The special value 0xFFFFFFFF can be used to indicate the same label is to be valid across all component links.

For types 8, 9, 25 and 26 the length field is variable and the Value field is a label as defined in [RFC3471]. As with all uses of labels, it is assumed that any node that can process the label information knows the syntax and semantics of the label from the context. Note that all TLVs are zero-padded to a multiple four octets so that if a label is not itself a multiple of four octets it must be disambiguated from the trailing zero pads by knowledge derived from the context.

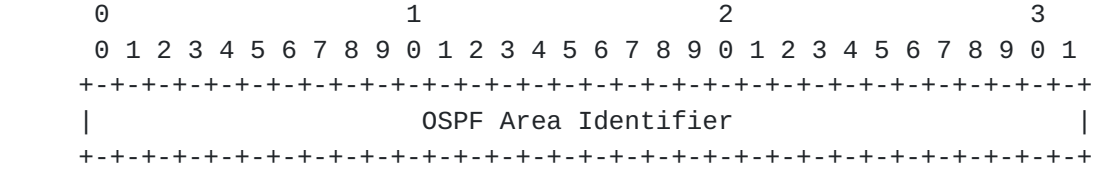
For types 10 and 27 the Value field has the format:



Router Id: 32 bits

The Router Id used to identify the node within the IGP.

For types 11 and 28 the Value field has the format:

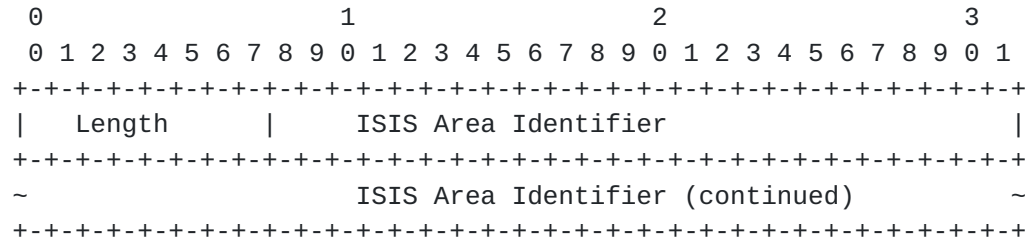


OSPF Area Identifier

The 4-octet area identifier the node is part of. In the case of

ABRs, this identifies the area where the failure has occurred.

For types 12 and 29 the Value field has the format:



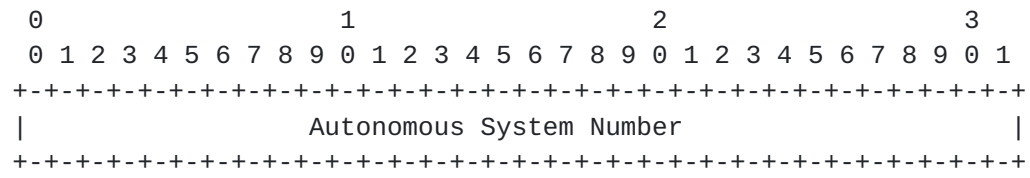
Length

Length of the actual (non-padded) ISIS Area Identifier in octets. Valid values are from 2 to 11 inclusive.

ISIS Area Identifier

The variable-length ISIS area identifier. Padded with trailing zeroes to a four-octet boundary.

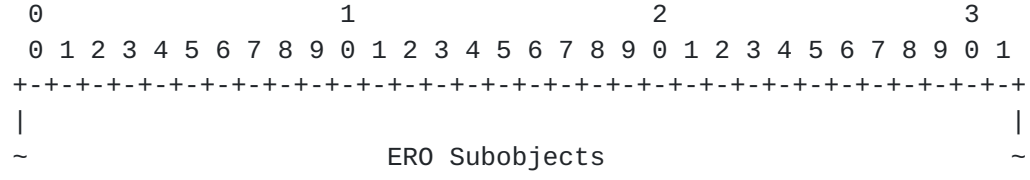
For types 13 and 30 the Value field has the format:

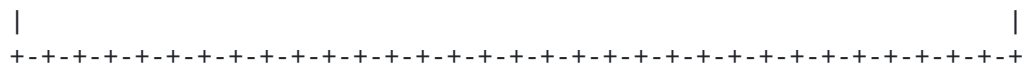


Autonomous System Number: 32 bits

The AS Number of the associated Autonomous System. Note that if 16-bit AS numbers are in use, the low order bits (16 through 31) should be used and the high order bits (0 through 15) should be set to zero.

For types 14, 15 and 31 the Value field has the format:

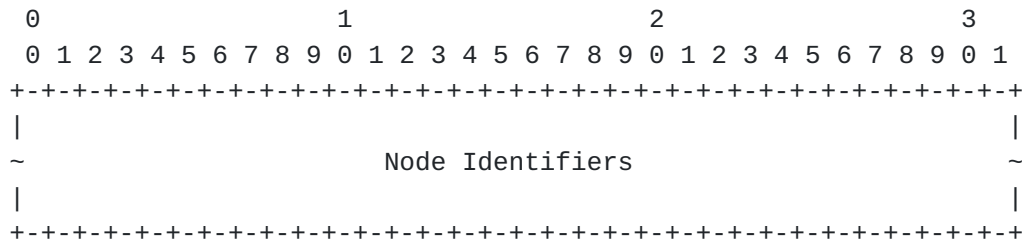




ERO Subobjects:

A sequence of ERO subobjects. Any ERO subobjects are allowed whether defined in [RFC3209], [RFC3473] or other documents. Note that ERO subobjects contain their own type and length fields.

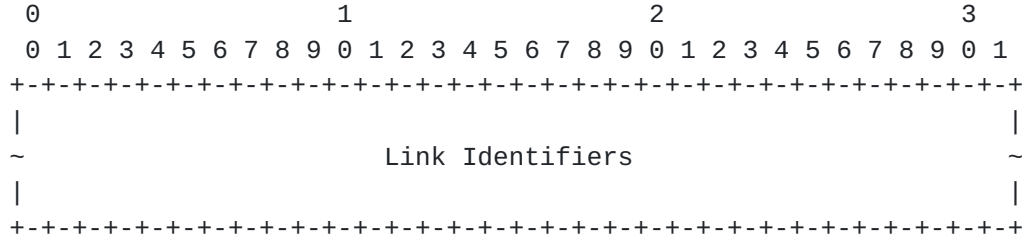
For type 32 the Value field has the format:



Node Identifiers:

A sequence of TLVs as defined here of types 1, 2 or 10 that indicates downstream nodes that have already participated in crankback attempts and have been declared unusable for the current LSP setup attempt.

For type 33 the Value field has the format:



Link Identifiers:

A sequence of TLVs as defined here of types 3, 4, 5, 6 or 7 that indicates incoming interfaces at downstream nodes that have already participated in crankback attempts and have

been declared unusable for the current LSP setup attempt.

7.2.1 Guidance for Use of IF_ID Error Spec TLVs

If Crankback is not being used but an IF-ID Error_Spec Object is included in a PathErr, ResvErr or Notify message, the sender SHOULD include one of the TLVs of type 1 through 5 as described in [RFC3473]. A sender that wishes to report an error with a component link of an unnumbered bundle SHOULD use the new TLVs of type 6 or 7 as defined in this document. A sender MAY include additional TLVs from the range 8 through 33 to report crankback information, although this information will at most only be used for logging.

If Cranback is being used, the sender of a PathErr, ResvErr or Notify message MUST use the IF_ID Error_Spec Object and MUST include at least one of the TLVs in the range 1 through 7 as described in [RFC3473] and the previous paragraph. Additional TLVs SHOULD also be included to report further information. Note that all such TLVs are optional and MAY be omitted. Inclusion of the optional TLVs SHOULD be performed where doing so helps to facilitate error reporting and crankback. The TLVs fall into three categories: those that are essential

to report the error, those that provide additional information that is or may be fundamental to the utility of cranback, and those that provide additional information that may be useful for crankback in some circumstances.

Many of the TLVs report the specific resource that has failed. For example, TLV type 1 can be used to report that the setup attempt was blocked by some form of resource failure on a specific interface identified by the IP address supplied. TLVs in this category are 1 through 13. These TLVs SHOULD be supplied whenever the node detecting and reporting the failure with crankback information has the information available. The use of TLVs of type 10, 11, 12 and 13, MAY, however, be omitted according to local policy and relevance of the information.

Reporting nodes SHOULD also supply TLVs from the range 14 through 26 as appropriate for reporting the error. The reporting nodes MAY also supply TLVs from the range 27 through 33.

Note that in deciding whether a TLV in the range 14 through 26 "is appropriate", the reporting node should consider amongst other things, whether the information is pertinent to the cause of the failure. For example, when a cross-connection fails it may be that the outgoing interface is faulted, in which case only the interface (for example, TLV type 1) needs to be reported, but if the problem is that the incoming interface cannot be connected to the outgoing interface because of temporary or permanent cross-connect limitations, the node should also include reference to the incoming interface (for example, TLV type 18).

Some TLVs help to locate the fault within the context of the path of the LSP that was being set up. TLVs of types 14, 15, 16 and 17 help to set the context of the error within the scope of an explicit path that has loose hops or non-precise abstract nodes. The ERO context information is not always a requirement, but a node may notice that it is a member of the next hop in the ERO (such as a loose or non-specific abstract node) and deduce that its upstream neighbor may have selected the path using next hop routing. In this case, providing the ERO context will be useful to the node further that performs re-routing.

Four TLVs (27, 28, 29 and 30) allow the location of the reporting node to be expanded upon. These TLVs would not be included if the information is not of use within the local system, but might be added by ABRs relaying the error. Note that the Reporting Node Id (TLV 27) need not be included if the IP address of the reporting node as indicated in the Error Spec itself, is sufficient to fully identify the node.

The last three TLVs (31, 32, and 33) provide additional information for recomputation points. The reporting node (or some node forwarding the error) may supply suggestions about the ERO that could have been used to avoid the error. As the error propagates back upstream and as crankback routing is attempted and fails, it is beneficial to collect lists of failed nodes and links so that they will not be included in further computations performed at upstream nodes. Theses lists may also be factored into route exclusions [EXCLUDE].

Note that there is no ordering requirement on any of the TLVs within the IF_ID Error Spec, and no implication should be drawn from the ordering of the TLVs in a received IF_ID Error Spec.

It is left as an implementation detail precisely when to include each of the TLVs according to the capabilities of the system reporting the error.

7.2.2 Alternate Path identification

No new object is used to distinguish between Path/Resv messages for an alternate LSP. Thus, the alternate LSP uses the same SESSION and SENDER_TEMPLATE/FILTER_SPEC objects as the ones used for the initial LSP under re-routing.

7.3. Action on Receiving Crankback Information

7.3.1 Re-route Attempts

As described in section 3, a node receiving crankback information in a PathErr must first check to see whether it is allowed to perform re-routing. This is indicated by the Re-routing Flags in the SESSION_ATTRIBUTE object during LSP setup request.

If a node is not allowed to perform re-routing it should forward the PathErr message, or if it is the ingress report the LSP as having failed.

If re-routing is allowed, the node should attempt to compute a path to the destination using the original (received) explicit path and excluding the failed/blocked node/link. The new path should be added to an LSP setup request as an explicit route and signaled.

LSRs performing crankback re-routing should store all received crankback information for an LSP until the LSP is successfully established or until the node abandons its attempts to re-route the LSP. This allows the combination of crankback information from multiple failures when computing an alternate path.

It is an implementation decision whether the crankback information is discarded immediately upon successful LSP establishment or retained for a period in case the LSP fails.

7.3.2 Location Identifiers of Blocked Links or Nodes

In order to compute an alternate path by crankback re-routing, it is necessary to identify the blocked links or nodes and their locations. The common identifier of each link or node in an MPLS network should be specified. Both protocol-independent and protocol- dependent identifiers may be specified. Although a general identifier that is independent of other protocols is preferable, there are a couple of restrictions on its use as described in the following subsection.

In link state protocols such as OSPF and IS-IS , each link and node in a network can be uniquely identified. For example, by the context of a Router ID and the Link ID. If the topology and resource information obtained by OSPF advertisements is used to compute a constraint-based path, the location of a blockage can be represented by such identifiers.

Note that, when the routing-protocol-specific link identifiers are used, the Re-routing Flag on the LSP setup request must have been set to show support for boundary or segment-based re-routing.

In this document, we specify routing protocol specific link and node identifiers for OSPFv2 for IPv4, IS-IS for IPv4, OSPF for IPv6, and IS-IS for IPv6. These identifiers may only be used if segment-based re-routing is supported, as indicated by the Routing Behavior flag on the LSP setup request.

7.3.3 Locating Errors within Loose or Abstract Nodes

The explicit route on the original LSP setup request may contain a loose or an Abstract Node. In these cases, the crankback information may refer to links or nodes that were not in the original explicit route.

In order to compute a new path, the repair point may need to identify the pair of hops (or nodes) in the explicit route between which the error/blockage occurred.

To assist this, the crankback information reports the top two hops of the explicit route as received at the reporting node. The first hop will likely identify the node or the link, the second hop will identify a 'next' hop from the original explicit route.

7.3.4 When Re-routing Fails

When a node cannot or chooses not to perform crankback re-

routing it must forward the PathErr message further upstream.

However, when a node was responsible for expanding or replacing the explicit route as the LSP setup was processed it MUST update the crankback information with

regard to the explicit route that it received. Only if this is done will the upstream nodes stand a chance of successfully routing around the problem.

7.3.5 Aggregation of Crankback Information

When a setup blocking error or an error in an established LSP occurs and cranback information is sent in an error notification message, some node upstream may choose to attempt crankback re-routing. If that node's attempts at re-routing fail the node will accumulate a set of failure information. When the node gives up it must propagate the failure message further upstream and include crankback information when it does so.

There is not scope in the protocol extensions described in this document to supply a full list of all of the failures that have occurred. Such a list would be indefinitely long and would include more detail than is required. However, TLVs 32 and 33 allow lists of unusable links and nodes to be accumulated as the failure is passed back upstream.

Aggregation may involve reporting all links from a node as unusable by flagging the node as unusable, or flagging an ABR as unusable when there is no downstream path available, and so on. The precise details of how aggregation of crankback information is performed are beyond the scope of this document.

7.4. Notification of Errors

7.4.1 ResvErr Processing

As described above, the resource allocation failure for RSVP-TE may occur on the reverse path when the Resv message is being processed. In this case, it is still useful to return the received crankback information to the ingress LSR. However, when the egress LSR receives the ResvErr message, per RFC 2205 it still has the option of re-issuing the Resv with different resource requirements (although not on an alternate path).

When a ResvErr carrying crankback information is received at an egress LSR, the egress LSR MAY ignore this object and perform the same actions as for any other ResvErr. However, if the egress LSR supports the crankback extensions defined in this document, and after all local recovery procedures have failed, it SHOULD generate a PathErr message carrying the crankback information and send it to the ingress LSR.

If a ResvErr reports on more than one FILTER_SPEC (because the Resv carried more than one FILTER_SPEC) then only one set of crankback information should be present in the ResvErr and it should apply to all FILTER_SPEC carried. In this case, it may be necessary per [RFC 2205] to generate more than one PathErr.

7.4.2 Notify Message Processing

[RFC3473] defines the Notify message to enhance error reporting in RSVP-TE networks. This message is not intended to replace the PathErr and ResvErr messages. The Notify message is sent to addresses requested on the Path and Resv messages. These addresses could (but need not) identify the ingress and egress LSRs respectively.

When a network error occurs, such as the failure of link hardware, the LSRs that detect the error MAY send Notify messages to the requested addresses. The type of error that causes a Notify message to be sent is an implementation detail.

In the event of a failure, an LSR that supports [RFC3473] and the crankback extensions defined in this document MAY choose to send a Notify message carrying crankback information. This would ensure a speedier report of the error to the ingress/egress LSRs.

7.5. Error Values

Error values for the Error Code "Admission Control Failure" are defined in [RFC2205]. Error values for the error code "Routing Problem" are defined in [RFC 3209] and [RFC 3473].

A new error value is defined for the error code "Routing Problem". "Re-routing limit exceeded" indicates that re-routing has failed because the number of crankback re-routing attempts has gone beyond the predetermined

threshold at an individual LSR.

7.6. Backward Compatibility

It is recognized that not all nodes in an RSVP-TE network will support the extensions defined in this document. It is important that an LSR that does not support these extensions can continue to process a PathErr, ResvErr or Notify message even if it carries the newly defined IF_ID ERROR_SPEC information (TLVs).

8. Routing Protocol Interactions

If the routing-protocol-specific link or node identifiers are used in the Link and Node IF_ID ERROR_SPEC TLVs defined above, the signaling has to interact with the OSPF/IS-IS routing protocol.

For example, when an intermediate LSR issues a PathErr message, the signaling module of the intermediate LSR should interact with the routing logic to determine the routing-protocol-specific link or node ID where the blockage or fault occurred and carry this information onto the Link TLV and Node TLV inside the IF_ID ERROR_SPEC object. The ingress LSR, upon receiving the

error message, should interact with the routing logic to compute an alternate path by pruning the specified link ID or node ID in the routing database.

Procedures concerning these protocol interactions are out of scope of this document.

9. LSP Restoration Considerations

LSP restoration is performed to recover an established LSP when a failure occurs along the path. In the case of LSP restoration, the extensions for crankback re-routing explained above can be applied for improving performance. This section gives an example of applying the above extensions to LSP restoration. The goal of this example is to give a general overview of how this might work, and not to give a detailed procedure for LSP restoration.

Although there are several techniques for LSP restoration, this section explains the case of on-demand LSP restoration, which attempts to set up a new LSP on demand after detecting an LSP failure.

9.1. Upstream of the Fault

When an LSR detects a fault on an adjacent downstream link or node, a PathErr message is sent upstream. In GMPLS, the ERROR_SPEC object may carry a Path_State_Remove_Flag indication. Each LSR receiving the message then releases the corresponding LSP. (Note that if the state removal indication is not present on the PathErr message, the ingress node must issue a PathTear message to cause the resources to be released.) If the failed LSP has to be restored at an upstream LSR, the IF_ID ERROR SPEC that includes the location information of the failed link or node is included in the PathErr message. The ingress, intermediate area border LSR, or indeed any repair point permitted by the Re-routing Flags, that receives the PathErr message can terminate the message and then perform alternate routing.

In a flat network, when the ingress LSR receives the PathErr message with the IF_ID ERROR_SPEC TLVs, it computes an alternate path around the blocked link or node satisfying the QoS constraints. If an alternate path is found, a new Path message is sent over this path toward the egress LSR.

In a network segmented into areas, the following procedures can be used. As explained in Section 8.2, the LSP restoration behavior is indicated in the Flags field of the SESSION_ATTRIBUTE object of the Path message. If the Flags indicate "End-to-end re-routing", the PathErr message is returned all the way back to the ingress LSR, which may then issue a new Path message along another path, which is the same procedure as in the flat network case above.

If the Flags field indicates Boundary re-routing, the ingress area border LSR MAY terminate the PathErr message and then perform alternate routing within the area for which the area border LSR is the ingress LSR.

If the Flags field indicates segment-based re-routing, any node MAY apply the procedures described above for Boundary re-routing.

9.2. Downstream of the Fault

This section only applies to errors that occur after an

LSP has been established. Note that an LSR that generates a PathErr with Path_State_Remove Flag SHOULD also send a PathTear downstream to clean up the LSP.

A node that detects a fault and is downstream of the fault MAY send a PathErr or Notify message containing an IF_ID ERROR SPEC that includes the location information of the failed link or node, and MAY send a PathTear to clean up the LSP at all other downstream nodes. However, if the reservation style for the LSP is Shared Explicit (SE) the detecting LSR MAY choose not to send a PathTear - this leaves the downstream LSP state in place and facilitates make-before-break repair of the LSP re-utilizing downstream resources. Note that if the detecting node does not send a PathTear immediately then unused sate will timeout according to the normal rules of [RFC2205].

At a well-known merge point, an ABR or an ASBR, a similar decision might also be made so as to better facilitate make-before-break repair. In this case a received PathTear might be 'absorbed' and not propagated further downstream for an LSP that has SE reservation style. Note, however, that this is a divergence from the protocol and might severely impact normal tear-down of LSPs.

10. IANA Considerations

10.1 Error Codes

A new error value is defined for the RSVP-TE "Routing Problem" error code that is defined in [RFC3209].

TBD Re-routing limit exceeded.

10.2 IF_ID_ERROR_SPEC TLVs

Note that the IF_ID_ERROR_SPEC TLV type values are not currently tracked by IANA. This might be a good opportunity to move them under IANA control.

10.3 LSP_ATTRIBUTES Object

Three bits are defined for inclusion in the LSP Attributes TLV of the LSP_ATTRIBUTES object. IANA is requested to assign those bits.

11. Security Considerations

It should be noted that while the extensions in this document introduce no new security holes in the protocols, should a malicious user gain protocol access to the network, the crankback information might be used to prevent establishment of valid LSPs.

The implementation of re-routing attempt thresholds are particularly important in this context.

The crankback routing extensions and procedures for LSP restoration as applied to RSVP-TE introduce no further new security considerations. Refer to [RFC2205], [RFC3209] and [RFC3473] for a description of applicable security considerations.

12. Acknowledgments

We would like to thank Juha Heinanen and Srinivas Makam for their review and comments, and Zhi-Wei Lin for his considered opinions. Thanks, too, to John Drake for encouraging us to resurrect this document and consider the use of the IF-ID ERROR SPEC object. Thanks for a welcome and very thorough review by Dimitri Papadimitriou.

13. Intellectual Property Considerations

14. Normative References

[RFC2205] R. Braden, et al., "Resource ReSerVation Protocol (RSVP) Version 1 Functional Specification", RFC2205, September 1997.

[RFC3209] D. Awduche, et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.

[RFC3471] P. Ashwood-Smith and L. Berger, et al., "Generalized MPLS - Signaling Functional Description", RFC 3471, January 2003.

[RFC3473] L. Berger, et al., "Generalized MPLS Signaling - RSVP-TE Extensions", RFC 3473, January 2003.

[LSP-ATTRIB] A. Farrel, D. Papadimitriou, JP. Vasseur, "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using RSVP-TE", draft-farrel-mpls-rsvpte-attributes-00.txt, October 2003, work in progress.

[ASON-REQ] D. Papadimitriou, J. Drake, J. Ash, A. Farrel, L. Ong, "Requirements for Generalized MPLS (GMPLS) Signaling Usage and Extensions for Automatically Switched Optical Network (ASON)", daft-ietf-ccamp-gmpls-ason-reqts-03.txt October 2003, work in progress.

15. Informational References

[ASH1] G. Ash, ITU-T Recommendations E.360.1 --> E.360.7, "QoS Routing & Related Traffic Engineering Methods for IP-, ATM-, & TDM-Based Multiservice Networks", May, 2002.

[FASTRR] Ping Pan, et al., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", draft-ietf-mpls-rsvp-lsp-fastreroute-03.txt, July 2003 (work in progress).

[G8080] ITU-T Recommendation G.808/Y.1304, Architecture for the Automatically Switched Optical Network (ASON), November 2001.

[EXCLUDE] C-Y. Lee, A. Farrel and S De Cnodder, "Exclude Routes - Extension to RSVP-TE", draft-ietf-ccamp-rsvp-te-exclude-route-00.txt, June 2003 (work in progress).

[PNNI] ATM Forum, "Private Network-Network Interface Specification Version 1.0 (PNNI 1.0)", <af-pnni-0055.000>, May 1996.

[RFC2702] D. Awduche, et al., "Requirements for Traffic Engineering Over MPLS", RFC2702, September 1999.

[RFC3469] V. Sharma, et al., "Framework for MPLS-based Recovery", RFC 3469, February 2003.

[INTER-AS] JP. Vasseur, and R. Zhang, "Inter-AS MPLS Traffic Engineering", draft-vasseur-inter-as-te-01.txt, June 2003, work in progress.

16. Authors' Addresses

Adrian Farrel (editor)
Old Dog Consulting
Phone: +44 (0) 1978 860944
EMail: adrian@olddog.co.uk

Arun Satyanarayana
Movaz Networks, Inc.
7926 Jones Branch Drive, Suite 615
McLean, VA 22102
Phone: (+1) 703-847-1785
EMail: aruns@movaz.com

Atsushi Iwata
NEC Corporation
Networking Research Laboratories
1-1, Miyazaki, 4-Chome, Miyamae-ku,
Kawasaki, Kanagawa, 216-8555, JAPAN
Phone: +81-(44)-856-2123
Fax: +81-(44)-856-2230
EMail: a-iwata@ah.jp.nec.com

Norihito Fujita
NEC Corporation
Networking Research Laboratories
1-1, Miyazaki, 4-Chome, Miyamae-ku,
Kawasaki, Kanagawa, 216-8555, JAPAN
Phone: +81-(44)-856-2123
Fax: +81-(44)-856-2230
EMail: n-fujita@bk.jp.nec.com

Gerald R. Ash
AT&T
Room MT D5-2A01
200 Laurel Avenue
Middletown, NJ 07748, USA
Phone: (+1) 732-420-4578
Fax: (+1) 732-368-8659
EMail: gash@att.com

Simon Marshall-Unitt
Data Connection Ltd.
100 Church Street
Enfield, Middlesex
EN2 6BQ, UK
Phone: (+44) (0)-208-366-1177
EMail: smu@dataconnection.com

17. Full Copyright Statement