Crankback Routing Extensions for CR-LDP

Abstract

This draft proposes crankback routing extensions for CR-LDP signaling. Recently, several routing protocol extensions for advertising resource information in addition to topology information have been proposed for use in distributed constraint-based routing. In such a distributed routing environment, however, the information used to compute a constraint-based path may be out of date. This means that label requests may be blocked by links or nodes without sufficient resources. This draft specifies crankback routing extensions for CR-LDP so that the label request can be retried on an alternate path that detours around the blocked link or node upon a setup failure. Furthermore, the proposed crankback routing schemes can be also applied to end-to-end LSP restoration by indicating the
location of the failure link or node. This would significantly improve the recovery ratio for failed LSPs, especially in situations where a large number of setup requests are triggered at the same time.

1. Introduction

CR-LDP (Constraint-based Routing Label Distribution Protocol) is defined in [CR-LDP] for establishing explicitly routed LSPs (CR-LSPs) in an MPLS network. Using CR-LDP, resources can also be reserved along a path to guarantee or control QoS for traffic carried on the CR-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 [OSPF], can be extended to distribute additional state information [LI]. Explicit paths can be designated based on the distributed information at the LSR initiating a CR-LSP and, if necessary, intermediate gateway 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In a distributed routing environment, however, the resource information used to compute a constraint-based path may be out of date.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. When a setup failure occurs, a notification is returned to the setup initiator (ingress LSR). In the current CR-LDP, the ingress LSR receiving the notification has to terminate the message and give up the LSP establishment.

If the ingress or intermediate gateway LSR knows the location of the blocked link or node, the LSR can designate an alternate path and then reissue the setup request, which can be achieved by the mechanism known as crankback routing [PNNI, ASH]. We propose the use of crankback routing in CR-LDP. Crankback routing requires notifying an upstream LSR of the location of the blocked link or node.

On the other hand, various restoration schemes for link or node failures have been proposed in [SWALLOW, MAKAM, SHARMA] and others. Fast restoration by pre-establishing a backup LSP is useful for failures on a primary LSP. If both the primary and backup paths fail, however, it is necessary to repair the LSP on an end-to-end basis. End-to-end restoration for alternative routing requires the location of the failed link or node. A proposed crankback routing scheme 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. The LSP restoration must therefore be performed with inaccurate information, which is likely to cause setup blocking.
Crankback routing would also improve failure recovery in these situations.

Recently, Multi-Protocol Lambda Switching has also been discussed. 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 OXC on an end-to-end explicit path. Furthermore, end-to-end restoration is the only way to recover LSP failures [CHAUDHURI]. This implies that crankback routing would also be useful in an MPLambdaS network. This draft proposes a crankback routing system that is an extension of CR-LDP and discusses the identification of blocked links or nodes in an MPLS network.

Although we address CR-LDP as the label distribution technique in the following, a similar discussion can be applied to TE-RSVP [TE-RSVP].

2. Explicit versus implicit rerouting indications

There have been problems in service provider networks when "inferring" from indirect information that rerouting is allowed. In the case of using an explicit rerouting indication, rerouting is explicitly authorized and not inferred. In the feedback case [ASHW], or in the current notification error messages [CR-LDP], one can infer a situation where rerouting can be done, but it also can lead to other problems, which have been experienced in practice, as illustrated below.

```
*-----------------*  *-----------------*
\|\                  |  |                 |
N2 ----------- N3- ------- AT--- EO2
\|\                  |  \|\ x-- --|--x x \\
\|\                  \|\ x-- x
\|\                  \|\ x-- --|--x x \\
N1 ----------- N4- ------- EO1
*-----------------*  *-----------------*
\|\    AS-1        |  |      AS-2    |
```

Figure 1. Example of network topology

Experiences of using release messages in TDM-based networks for analogous purposes provide some guidance. One can use a release message with a cause value (CV) indicating "link congestion" (a CV already standardized in ISUP, for example) to try to then do rerouting at the originating node. However, this sometimes leads to other problems. Figure 1 is used to illustrate five examples based on actual service-provider experiences with respect to crankback (i.e., explicit indication) versus release/CV, or "no bandwidth
available" (NBA) (i.e., implicit indication) processing. In this example, N1, N2, N3, and N4 are located in one area (AS-1), and AT, EO1, and EO2 are in another area (AS-2).

(1) A connection request from node N1 to EO1 may route to N4 and then find "all circuits busy" (equivalent to NBA). N4 returns a release message to N1 with cause value (CV) 34 (indicates all circuits busy/NBA). 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 then it will 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) Now 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, however N2 will not realize where this blocking occurred based on the CV34, and in this case there is no point in further alternate routing. However with RA it may try to route N2-N1-N4-AT-EO2, but of course this fails again. With crankback, however, this could be resolved, since crankback also indicates where the blocking occurred, and in this scenario a crankback should not be returned. Rather, a CV34 should be returned and should result in blocking the connection request at N2 (the proper response to CV34).

(3) However in another case of a connection request from N2 to EO2, suppose that link N3-AT is blocked, then in this case N3 should return a crankback (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, given that EO1-N4 is already known blocked due to NBA) rather than trying to alternate route through AT-N3-N2, which may well be successful. Had N3 returned a crankback, the EO1 could respond by trying the alternate route.

(5) Granted that with topology exchange, such as OSPF, the ingress LSR could infer the rerouting condition. However, one of the reasons for crankback is to avoid the overhead or available-link-bandwidth (ALB) flooding to more efficiently use local state information to
direct alternate routing at the ingress-LSR. The draft [ASH2] shows how event-dependent-routing (EDR) can just use crankback, and not ALB flooding as required by state-dependent-routing (SDR), to decide on the path in the network through "learning models". Reducing the ALB flooding reduces overhead and can lead to large AS size.

Therefore, the alternate routing should be indicated based on an explicit indication (as in examples 2 and 5), and it is best to know the following information separately:

a) where blockage/congestion occurred (as in examples 2-3), and

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

3. Crankback Routing Behaviors

Crankback routing can be performed in the following situation.

a. A setup request is blocked.

b. An established LSP fails.

Although crankback routing can be done on a hop-by-hop basis, which may be useful for fast restoration, we address source-route based crankback routing in this draft.

3.1. Crankback Routing due to Setup Request Blockage

When a label request is blocked due to unavailable resources, a notification message with a Rerouting TLV with the location identifier of the blockage, is returned to the LSR initiating the label request (ingress LSR). The notification message is called a "crankback notification". It may include information such as bandwidth availability, as described in [ASHW]. This issue is left for further study.

In a flat network without segmentation, the ingress LSR receives the message and designates 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 label request is sent over this path. In a network segmented into areas such as hierarchical OSPF, the area border node may terminate the crankback notification and then perform alternate routing instead of the ingress LSR.

The LSR that computed the alternate path stores the location
identifiers of the blockages indicated in a crankback notification until it receives a corresponding label mapping message. Upon receiving a crankback notification again after retrying a label request, another alternate path is computed based on the previous blocked links or nodes in the stored information as well as the current ones in the received notification message.

Optionally, the maximum number of crankback routing allowed for establishing an LSP may be limited. This is useful to prevent a endless repetition of crankback routing in certain error conditions. It is also useful for reducing the number of unnecessary retries, which do not improve performance. When crankback notifications exceed the threshold, the message that exceeded the threshold is processed as an ordinary (non-crankback) notification and no more label requests are issued. Other mechanisms for controlling crankback routing can also be used.

3.2. Crankback Routing due to LSP Failures

When an LSR detects a fault on an adjacent downstream link or node, a label withdraw message is sent upstream over a CR-LSP configured on the LSR. Each LSR receiving the message then releases the corresponding LSP. If the failed LSP is expected to be restored at an upstream LSR, the Rerouting TLV with the location information of the fault link or node is included in a label withdraw message. The ingress or intermediate gateway LSR receiving the message can terminate this messages and then perform alternate routing.

On the downstream side of the failure, an LSR detecting the failure sends a label release message over a CR-LSP configured on the LSR. The egress or intermediate gateway LSR receiving the message can terminate this message and wait until the new label request for restoration is received.

Other behaviors are the same as those of performing crankback routing upon setup request blockage.

4. Explicit indication code of rerouting

Crankback rerouting action is indicated in the Status TLV in the Notification message [CR-LDP]. Following is a new status code for the Status TLV. The exact value of YY will be defined. When an ingress LSR receiving this Status TLV does not support the rerouting functions explained above, this TLV is allowed to be silently ignored.

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>--------------</td>
<td>------</td>
</tr>
</tbody>
</table>
5. Location Identifiers of Blocked Links or Nodes

In order to compute an alternate path by crankback routing, it is necessary to identify where blocked links or nodes are. The common identifier of each link or node in an MPLS network should be specified. We propose both protocol-independent and protocol-dependent identifiers. Although a general identifier that is independent of other protocols is preferable, there are a couple of restrictions on its use.

In a CR-LDP label request, an explicit route can be represented as a list of ER-Hop TLVs in the ER-TLV. The first ER-Hop TLV is removed when the label request is sent to the next abstract node. Therefore, we propose that the blockage location be represented using the top ER-Hop TLVs upon a blockage. In the case of a node blockage, the first ER-Hop TLV is returned in a crankback notification. Upon a link blockage, the first and second ER-Hop TLVs is returned, which means that the blockage location is the link between the two nodes indicated by these TLVs. If the abstract node described by the ER-Hop TLV is represented by a strict hop with a prefix length of 32, this indicates one IP address of a single IPv4 node. If this condition is met, the indication of the blockage location by the ER-Hop TLVs allows an ingress or intermediate gateway LSR to compute an alternate path to detour just the link/node. If not, the link or node where a setup failure occurred probably will not be identified. Furthermore, if an initial label request is traversed with hop-by-hop routing, the label request message does not have an ER-TLV. These examples imply that ER-Hop TLVs cannot always be used as the location identifier of a blockage. An alternate indicator of such blocked links or nodes is required.

In link state protocols such as OSPF and IS-IS, each link and node in a network can be uniquely identified. For example, the OSPF protocol uses Router ID as the node identifier and the combination of Router ID and Link ID as the link identifier. 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. We also propose protocol-specific link or node identifiers that can be used for multiple routing protocols. In this draft, we specify identifications for OSPF and IS-IS. Extension for other routing protocols, such as BGP, etc., can be easily applied but is left for further study.

Although a general link/node identifier in a network is preferable, it is difficult to specify it because MPLS signaling protocols are...
independent of routing protocols. This issue is left for further study.

6. Additional TLVs

In this section, we define additional TLVs included in notification and label withdraw messages for indicating the location of blocked links or blocked nodes.

6.1. Rerouting TLV

When resource allocation for a label request is not allowed, a notification message is returned. In a network where crankback routing is supported, the Rerouting TLV is included in the notification message. This message must carry the LSPID TLV of the corresponding CR-LSP. The Status TLV included in the notification message indicates the cause of the problem. Additional status codes can be added to the current defined ones to give more detail about the blockage such as "Reserved" and "Heavily Loaded" [ASH].

If LSP restoration is supported, the Rerouting TLV is included in label withdraw messages caused by network failures. The message must also carry the LSPID TLV of the corresponding CR-LSP. The Status TLV can be also included to indicate the cause of the problem.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|     Rerouting (TBD)       |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Location Identifier Components                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Location Identifier Components

One or more TLVs of the following are included.

- Link TLV: See below for the format.
- Node TLV: See below for the format.

6.2. Link TLV

The Link TLV is used to identify a link where a setup blockage or a fault is detected.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
|0|0|       Link (TBD)          |            Length             |
|+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Reserved             | Protocol Type |     Num       |
|+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Link Identifier Components 1                  |
|+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         ............                          |
|                 Link Identifier Components N                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

**Protocol Type**

A one-octet unsigned integer containing the unique identifier of a protocol used for QoS routing. The following type numbers are defined. If a constraint-based route is to be calculated with no routing protocols, type 1 should be used. A location is identified using ER-Hop TLVs, independent of protocols.

1: CR-LDP (ER-Hop TLVs)
2: OSPF
3: IS-IS
128-255: Reserved for private and experimental use.

**Num; Number of Link Identifier Components**

A one-octet unsigned integer specifying the number of Link Identifier Components included in the TLV.

**One or more Link Identifier Components**

A variable length field containing link identifiers for relevant protocol types.

When the type is CR-LDP (1), the following format is used. The first and second ER-Hop TLVs upon setup blockage are included.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       First ER-Hop TLV                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Second ER-Hop TLV                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

When the protocol type is OSPF (2), the following 8-octet format is used.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
Router ID
The same value as the Router ID used in OSPF.

Link ID
The same value as the Link ID used in OSPF.

When the protocol type is IS-IS (3), the following 12-octet format is used.

```
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      +--------------------------------------------------+
      | System ID                                      |
      | +---------------+------------------|
      | Reserved       | Protocol Type    |
      | IP Interface Address |              |
      +--------------------------------------------------+
```

System ID
The same value as System ID used in IS-IS.

IP Interface Address
The IP address assigned to the interface advertised in IS-IS.

6.3. Node TLV
The Node TLV is used to identify a node where a setup blockage or a fault is detected.
Protocol Type
A one-octet unsigned integer containing the unique identifier of
the protocol used for QoS routing. The same value as defined in the
Link TLV is used.

Num; Number of Node Identifier Components
A one-octet unsigned integer specifying the number of Node
Identifier Components included in the TLV.

One or more Node Identifier Components
A variable length field containing the node identifier for relevant
protocol types.

When the type is CR-LDP (1), the following format is used. The
first ER-Hop TLV upon setup blockage is included.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       First ER-Hop TLV                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

When the protocol type is OSPF (2), the following 4-octet format is
used.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Router ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Router ID
The same value as the Router ID used in OSPF.

When the protocol type is IS-IS (3), the following 8-octet format
is used.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
| System ID | Reserved                                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

System ID
The same value as the System ID used in IS-IS.
7. Security Considerations

Security considerations are not addressed in this version of the draft.

8. Acknowledgments

We would like to thank Juha Heinanen and Srinivas Makam for their review and comments.

9. References


9. Authors’ Addresses

Atsushi Iwata
NEC Corporation
Computer & Communication Media Research
1-1, Miyazaki, 4-Chome, Miyamae-ku,
Kawasaki, Kanagawa, 216-8555, JAPAN

Phone: +81 (44) 856-2123
Fax: +81 (44) 856-2230
Email: iwata@ccm.cl.nec.co.jp

Norihito Fujita
NEC Corporation
Computer & Communication Media Research
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@ccm.cl.nec.co.jp

Gerald R. Ash
AT&T
Room MT E3-3C37
200 Laurel Avenue
Middletown, NJ 07748, USA

Phone: +1 732 420-4578
Fax: +1 732 440-6687
Email: gash@att.com