Abstract

Consider a BGP free core scenario with LDP running in the core. Suppose the edge BGP speakers PE1 and PE2 know about a prefix P/m via the external routers CE1 and CE2. If the edge LSR PE1 crashes or becomes totally disconnected from the core, it is desirable for a core LSR "P", that is carrying traffic to the failed edge LSR PE1, to immediately restore traffic by re-routing packets destined to the prefix P/m from the LSP terminating on PE1 to be forwarded over the LSP terminating on PE2, until BGP reconverges. If the packets originally flowing to the failed edge LSR PE1 are BGP labeled, then the repairing core LSR P must swap the label (corresponding to prefix P/m) advertised by the failed edge LSR PE1 with the label advertised for the same prefix by the edge LSR PE2 before re-routing the packets through an LSP terminating on PE2. To implement BGP edge node protection in a BGP-free LDP core, this document proposes an extension to LDP. This extension allows an LDP speaker running on an Edge LSR Node (e.g. PE1) to inform the LDP speakers running on core LSRs about the "Repair" edge LSR (e.g. PE2), as well as Repair LSR’s label for prefix P/m, if any.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

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/1id-abstracts.txt

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

This Internet-Draft will expire on September 5, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction.......................................................3
   1.1. Conventions used in this document............................4
   1.2. Terminology.................................................4
   1.3. Problem definition........................................5
2. The Proposed LDP Extension......................................5
   2.1. The LDP "BGP Repair Path status" Code......................5
   2.2. The LDP "BGP Repair Path Status" TLV.......................6
   2.3. LDP Capability Negotiation................................8
   2.4. BGP Repair Status in a LDP Notification message...........8
3. BGP Repair Path Signaling Procedures...........................9
   3.1. Signaling a BGP Repair Path.....................................9
   3.2. Updating a BGP Repair Path...................................10
   3.3. Withdrawing a BGP Repair Path.................................10
1. Introduction

In a BGP free core, where traffic is tunneled between edge routers/LSRs, (MP)BGP [2][3] speakers advertise reachability information about prefixes to edge routers only. For labeled address families, namely AFI/SAFI 1/4, 2/4, 1/128, and 2/128, an edge LSR assigns local labels to prefixes and associates the local label with each advertised prefix such as L3VPN [9], 6PE [10], and Softwire [8]. Suppose that a given edge LSR is chosen as the best next-hop for a prefix P/m. An ingress LSR receives a packet destined for the prefix P/m from an external router, and sends the packet to the egress LSR through an LSP terminating on the egress LSR. If the prefix P/m is a BGP labeled prefix, the ingress LSR pushes the BGP label advertised by the egress LSR before sending the packet into the LDP LSP terminating on the egress LSR. Upon receiving the packet from the core, the egress LSR takes the appropriate forwarding decision based on the content of the packet and/or the label(s) pushed on the packet.

In modern networks with redundancy in place, it is not uncommon to have a prefix reachable via multiple edge routers. One example is the best external path [7]. Another more common and widely deployed scenario is L3VPN [9] with multi-homed VPN sites. As an example, consider the L3VPN topology depicted in Figure 1.
As illustrated in Figure 1, the edge router PE0 is the primary NH for both 10.0.0.0/8 and 20.0.0.0/8. At the same time, both 10.0.0.0/8 and 20.0.0.0/8 are reachable through the other edge routers PE1 and PE2, respectively.

1.1. 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 [1].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

1.2. Terminology

- LSR : Label Switched Router (In the context of this document, this refers to a router doing label switching on LDP and/or BGP labels)

- LSP: Label Switched Path
1.3. Problem definition

The general problem for the example shown in Section 1. is specified in [6]. The objective of this document is to specify an LDP [4] extension to let the primary egress PE inform repairing core router(s) about the repair path in an LDP core for both labeled and unlabeled protected prefixes. In other words, this is an LDP-based implementation of step 3 in [6]. Other problems, such as determining the repair PE, detecting the protected PE (node/connectivity) failure, and interactions between LDP and BGP on protected edge PE LSR, are beyond the scope of this document.

2. The Proposed LDP Extension

As specified in [4] section 3.5.1, an LDP speaker can use LDP Notification message to send its status or advisory information towards a peer. An LDP Notification message consists of a Status TLV and optional parameters, whereas "Status" TLV holds the status code being signaled. For an egress PE LSR to convey repair path info (for a BGP next-hop) to core LSRs, we propose to convey this information via a LDP Notification message that carries a new status code in "LDP Status" TLV, a new "BGP Repair Path Status" TLV and FEC TLV (corresponding to BGP next-hop) as optional parameters. This information is to be advertised to a peer only if the peer has signaled the support for "Unrecognized Notification" capability a specified in [5].

The proposed extensions are described more in details in following sub-sections.

2.1. The LDP "BGP Repair Path status" Code

A new LDP status code, namely "BGP Repair Path status" is defined that is to be set in the "Status Code" field of the "Status TLV" as defined in [4] section 3.4.6.
2.2. The LDP "BGP Repair Path Status" TLV

A new LDP TLV, namely "BGP Repair Path Status TLV", is defined to be used in an LDP Notification message under optional parameters section only if the Notification message status code is set to "BGP Repair Path status".

This TLV is an implementation of the repair path defined in [6] and is used to convey the information about the Repair edge LSR and its associated BGP label, if any, for traffic destination prefix P/m. This information is conveyed in the context of the protected primary BGP nexthop [6], whose information is carried in the FEC TLV. This document allows only one repair path per BGP nexthop.

The encoding of the "BGP Repair Path Status TLV" is as follows:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F| Repair Path TLV Type(TBD) |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|A|L|P|     Reserved            |         Address Family        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Repair PE Address (variable size)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Underlying Repair label (optional)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2 Format of BGP Repair Path Status TLV

The fields are as follows

**U/F:**

Must be set to 1/0 respectively so that this TLV can be ignored if not known or not supported.

**Repair Path TLV Type:**

IANA assigned TLV value

**A bit**

Indicates if Repair Path information is to be "added" or "removed". MUST be set to 1 to signal addition of the information, and set to signal addition of the information, and set to

**L Bit:**

Expires September 5, 2012
Indicates whether optional "Underlying Repair Label" [6] field is present or not. Must be set to 1 if the TLV also contains/encodes "Underlying Repair Label", else must be set to 0. This bit MUST be set to 0 when A bit is set to 0.

P Bit:

If set, then the label in the "Underlying Repair label" sub field MUST be pushed instead of swapped. The "P" bit has the same semantics of the "Push" flag in [6]. If the "I" bit is zero then the "P" bit MUST be set to zero.

Reserved:

MUST be zero on transmit and ignored in receipt

Address Family

Identical to the "Address Family" field used in encoding the Prefix FEC element value specified in Section 3.4.1 in [4]. May be different of the address family of the prefix in the FEC

Repair PE Address

The length of this field is dependent on the Address Family field. This is either the 4 octet IPv4 address or 16 octet IPv6 address of a host address belonging to repair PE. The encoding of the Repair PE Address is identical to the encoding of the value of the IPv4 or IPv6 Transport Address TLV specified in Section 3.5.2 in [4]. This field encodes the "Repair Next-hop" defined in [6].

Underlying Repair Label (optional)

If included in the TLV, it is a single label defined in accordance to Generic Label TLV specified in Section 3.4.2.1 in [4]. This field encodes the "Underlying Repair Label" defined in [6].

Length

The length (in octets) of the TLV following the "Length" field. For example, the length MUST be 8 with IPv4 Repair PE address or 20 with IPv6 Repair PE address when L bit is set to 0, and MUST be 12 and 24 octets otherwise for IPv4 and IPv6 address families, respectively.
2.3. LDP Capability Negotiation

This BGP Repair Path information is to be computed by an edge PE LSR under a user configuration control. Once computed, this information can be unsolicitedly sent to core P LSRs for edge PE node protection, and it is up to the receiving P LSR to store and use this information to protect the edge PE LSR.

Given above procedures, no new LDP capability [RFC5561] negotiation is needed between PE and P LSRs to support this feature extensions. However, to ensure backward compatibility and deterministic behavior, it is required that this information be sent to only those P LSRs that support "Unrecognized Notification" capability as specified in [5]. This will ensure that these new status and TLV does not cause any issue at a receiving P LSR if not known or not supported, and be discarded in accordance with "Unrecognized Notification" procedures.

2.4. BGP Repair Status in a LDP Notification message

The general format of an LDP Notification message that carries information regarding BGP repair path is as follows:

```
| 0 | Notification (0x0001) | Message Length |
|++++++| Message ID |
|++++++| Status TLV |
|++++++| BGP Repair Path Status TLV |
|++++++| FEC TLV |
```

Figure 3 : LDP Notification message with BGP Repair Status

The "Status TLV" status code is set to "BGP Repair Path status" to indicate that the message is used to convey BGP Repair Path information. When this status code is set, a Notification message MUST contain both "BGP Repair Status TLV" and a "FEC TLV" in the message.

Since this notification does not refer to any particular message, the "Message ID" and "Message Type" fields in the "Status TLV" MUST be set to 0.
The "BGP Repair Path Status TLV" is encoded as described earlier.

The FEC TLV MUST contain a single "Prefix FEC Element" that encodes the BGP nexthop information as host prefix. This field encodes the "Primary Next-hop" defined in [6].

3. BGP Repair Path Signaling Procedures

To describe the signaling procedures clearly, let’s first assume that:

- Protected edge LSR is PE1, Repair edge LSR is PE2 and repairing core LSR is P1 LSR router, and P2 is also a core LSR that does not support BGP Repair Path functionality.
- IPv4 Host addresses for PE1 and PE2 are A1 and A2 respectively.
- Protected BGP IPv4 prefix is P/m.
- BGP label assigned by PE2 for P/m is L2.
- P1 and P2 LSR support LDP "Unrecognized Notification" Capability.

3.1. Signaling a BGP Repair Path

An operator enables this feature on PE1 using a configuration knob. The PE1 computes Repair PE information (PE2 address A2, and PE2 BGP label L2) for a given BGP prefix P/m. PE1 encodes the LDP Notification to advertise the Repair path information to all those core LSR peers (including P1/P2 LSRs) who have advertised "Unrecognized Notification" Capability TLV for given LDP session.

The PE1 LSR encodes the following TLVs:

- "Status" TLV: status code is set to "BGP Repair Path status" and "Message Type" and Message ID fields set to 0.
- "BGP Repair Path Status" TLV: This TLV is encoded with A-bit set to 1, L-bit set to 1, P-bit set to 0, Address Family set to 1 (IPv4), "Repair PE Address" field populated with A2, and "Repair PE’s Label" field set to L2.
- "FEC" TLV: This TLV is encoded with a single "Prefix FEC Element" whose Address Family is set to 1 (IPv4) to indicate IPV4 prefix, and prefix P/m encoded accordingly.

After encoding these TLVs, PE1 LSR bundles them in an LDP Notification message, as shown in Figure 3, and sends them to its upstream core peer P1 and P2 LSRs.
On receipt of this information, P1 stores this information and uses this to fast reroute the BGP destined traffic to PE2 upon PE2 node/connectivity failure detection. P2, on the other hand, does not recognize this new status in the LDP Notification message and hence discards it silently.

In order to be able to protect primary BGP nexthops, it is required that the repairing P LSR (P1) must have a LSP terminating on Repair PE host prefix (as indicated by "Repair PE Address" field in the received "BGP Repair Path Status" TLV).

3.2. Updating a BGP Repair Path

The repair path information is identified by

- the "primary next-hop" encoded in the "FEC TLV" shown in Figure 3 and,
- the "Repair Next-hop" encoded in the "Repair PE Address" shown in Figure 2.

Once a repair path has been signaled to P core LSR, it can be updated by simply sending another LDP Notification message using the procedures described in the previous section.

Upon receipt of a new repair path information, the LDP receiver (P1 LSR) MUST discard any previously learnt Repair information from the sending PE1 LSR, and update it with the most recently received.

3.3. Withdrawing a BGP Repair Path

Once a repair path has been signaled to P core LSR, it can be withdrawn by simply sending another LDP Notification message using the procedures described in the previous section with following changes:

- Set A-bit, L-bit, and P-bit in "BGP Repair Path Status" TLV to 0
- No "Underlying Repair Label" field included

Upon receipt of a withdrawal of a repair path, the LDP receiver (P1 LSR) MUST discard any previously learnt Repair information from the sending PE1 LSR for a given BGP prefix.

4. Security Considerations

No additional security risk is introduced by using the mechanisms proposed in this document
5. IANA Considerations

This document introduces the following new protocol elements that require code point assignment by IANA:

- "BGP Repair Path status" status code from "LDP Status Code Name Space" registry (requested code point: 0x00000050)
- "BGP Repair Path Status TLV" from "LDP TLV Type Name Space" registry (requested code point: 0x050F)

6. Conclusions

This document proposes an LDP extension that allows an egress PE to advertise a repair path consisting of a repair egress PE and an underlying label to repairing core router. Advertising this information to core routers allows core routers to provide FRR protection against primary egress PE node failure while keeping the core BGP-free.

7. References

7.1. Normative References


7.2. Informative References

Acknowledgments

Special thanks to Keyur Patel and Alton Lo for the valuable help.

This document was prepared using 2-Word-v2.0.template.dot.

Authors’ Addresses

Ahmed Bashandy
Cisco Systems
170 West Tasman Dr, San Jose, CA 95134
Email: bashandy@cisco.com

Kamran Raza
Cisco Systems,
2000 Innovation Dr., Kanata, ON K2K 3E8, Canada
EMail: skraza@cisco.com