Internet Engineering Task Force
INTERNET DRAFT

Olivier Bonaventure
Pierre Francois
UCLouvain
Mike Shand
Stefano Previdi
cisco

February, 2006
Expires August, 2006

ISIS extensions for ordered FIB updates
<draft-bonaventure-isis-ordered-00.txt>

Status of this Memo

Copyright Notice

Abstract

This document proposes extensions to allow IS-IS to support the ordered convergence defined in [RTG-OFIB] that allows to avoid transient forwarding loops during the FIB updates that follow a non-urgent topology change.

1 Introduction

In large ISP networks, topology changes are frequent events. When a link metric changes, the router responsible for the changing link floods a new LSP and forces all routers to recompute their routing table and update their FIB. This routing convergence is that it can lead to packet losses due to transient forwarding loops [RTG-OFIB].

There are various types of topology changes in IP networks. Sudden failures of unprotected links are urgent and should be handled by using a normal, fast, convergence of the link state routing protocol. However, there are frequent topology changes that are non-urgent. For example, operators often need to enable or disable links for short periods of time during maintenance operations. Another example are the IP networks where traffic engineering is performed by changing the ISIS link metrics. These events are non-urgent and they should not cause packet losses or transient forwarding loops. Transient forwarding loops can be avoided by forcing the routers to orderly update their FIB as proposed in [RTG-OFIB]

In this document, we propose IS-IS extensions allowing IS-IS routers to implement the ordered FIB update described in [RTG-OFIB]. We first describe the considered topology changes in section 2. Section 3 defines the new IS-IS TLVs that are used to support [RTG-OFIB].

2. Graceful topology changes

In this section, we describe two types of graceful topology changes. The first one is the change of a link metric (increase or decrease). The second one is the change of the settings of the overload bit (Set to Unset or Unset to Set).

2.1 Change in link metric

The first change that we consider is a modification of the metric associated to a link. This change can be a metric increase or a metric decrease. A non-urgent link shutdown can be converted into a weight change by doing the topology change in two steps. First, the metric of the link is set to the highest possible value. This triggers an ordered FIB update. After the FIB update, the link can be safely removed without risking transient loops. Similarly, a failure of a protected link [IPFRR] [MPLSFRR] can be handled as a non-urgent failure.

2.2 Overload status change

The second change that we consider is the transition of the overload

from Set to Unset or from Unset to Set.

3. Extensions to IS-IS

In this section, we describe the IS-IS extensions that are required to support ordered FIB updates that allow to avoid transient loops. There are two types of new TLVs. First, we define a capability sub-TLV to allow a router to advertise its ability to support ordered FIB updates. Second, we define a TLVs to be used inside the IS-IS Hello PDUs to implement the completion messages defined in [RTG-OFIB].

3.1 Ordered FIB capability sub-TLV

We define a FIB capability sub-TLV (advertised in the IS-IS CAPABILITY TLV defined in [ISIS-CAP]) to allow routers to determine which routers in the network support the ordered FIB updates. The FIB capability sub-TLV is composed of 1 octet for the type, 1 octet specifying the TLV length and a value field. The FIB capability TLV is used to advertise the type of ordered FIB updates supported by this router. The FIB capability sub-TLV shall only be flooded inside the current area.

    TYPE: TBD (Ordered FIB capability sub-TLV)
    LENGTH: 1 byte
    VALUE: Indicates the version of the ordered FIB update supported
by the router. This document defines version 1.

This ordered FIB capability sub-TLV must be sent in a router capability TLV where the S bit is set to 0x0 and the D bit is set to 0x0 according to the leaking rules defined in [ISISCAP].

3.2 Completion message TLV

We define a new type of TLV to encode inside the IS-IS Hello PDUs the completion messages proposed in [RTG-OFIB]. A IS-IS Hello PDU may carry several completion messages. Each completion message is encoded as a sub-TLV inside the completion message TLV.

    Type   TBD (Completion TLV)
    Length # of bytes in the value field (variable)
    Value : one or several sub-TLVs defined below

This document defines three types of sub-TLVs for the completion messages.

3.2.1 Metric change sub-TLV

The metric change sub-TLV shall be used as a completion message when the metric of a link changes. This happens when a metric is changed manually. It is also possible to use the mechanisms defined in [RTG-OFIB] and this document for link failures. For example, when a link is shutdown manually, the router could first advertise the link with metric maxmetric-1 as a non-urgent change to let all routers apply the ordered FIB update. Later, the link can be safely removed without risking transient loops. A router could use the same technique when one of its protected links fails.

We define two types of metric changes sub-TLV that are used inside the completion TLV. The first one is used with wide metrics and the second one with normal metrics. We expect that most deployments will use wide metrics.

```
Type   TBD (Wide Metric change sub-TLV)
Length # of bytes in the value field (21)
Value
                                    No. of bytes
             +-----------------------+
             | 0  0  0  0 0 0 0 |Ack |      1
             +-----------------------+
             | Upstream Node Id      |      7
             +-----------------------+
             | Downstream Node Id    |      7
             +-----------------------+
             | Old metric            |      3
             +-----------------------+
             | New metric            |      3
             +-----------------------+
```

Figure 1: New sub-TLV for (wide) metric change completion message

In the Wide metric change sub-TLV, the rightmost bit of the first byte is used to indicate whether it corresponds to a completion message (value 0) or an acknowledgment (value 1). The upstream and downstream node Ids are the system identifiers of the link whose metric changes. "Old metric" is the previous value of the wide metric for this link and "New metric" the current value.

A similar sub-TLV is also defined for the normal metric. As current implementations of IS-IS only support the default metric, we do not consider changes in the delay, error or expense metrics.

```
Type   TBD (Normal Metric change sub-TLV)
Length # of bytes in the value field (17)
Value
                                     No. of bytes
             +-----------------------+
```

```
                | 0  0  0  0 0 0 0 |Ack |     1
                +----------------------+
                | Upstream Node Id     |     7
                +----------------------+
                | Downstream Node Id   |     7
                +----------------------+
                | Old Default metric   |     1
                +----------------------+
                | New Default metric   |     1
                +----------------------+
```

Figure 2: New sub-TLV for (normal) metric change completion message

The Default metric shall be encoded as in normal LSPs.

3.2.2 Overload change sub-TLV

The overload change sub-TLV is used to encode the changes to the Overload Bit.

```
Type   TBD (Overload change sub-TLV)
Length # of octets in the value field (10)
Value
                                   No. of bytes

             +----------------------+
             |Old|New| 0 0 0 0 0 Ack |     1
             +----------------------+
             | Node Id              |     7
             +----------------------+
```

Figure 3: New sub-TLV for Overload bit change

The first octet contains three useful bits. The leftmost bit is the old value of the overload bit, the next bit is the new value of the overload bit and rightmost bit is the acknowledgment bit. The Node Id is the system identifier of the node whose overload bit change gracefully.

3.3 Utilization of the completion messages

The TLVs defined in sections 3.2 and 3.3 allow to implement the completion messages specified in [RTG-OFIB]. Each sub-TLV corresponds to a completion message for one topological change. The completion TLV in one HELLO PDU may carry several sub-TLVs and thus several completion messages. A router supporting the ordered FIB update

defined in this document SHOULD send HELLO PDUs containing the completion messages according to the rules defined in [RTG-OFIB].

To ensure a reliable delivery of the completion messages, an acknowledgment scheme is used. Upon reception of a HELLO PDU containing a completion message (rightmost bit of the first byte set to 0 in the sub-TLV), a router MUST ensure that the next HELLO PDU that it will send to this neighbor will contain, in the completion TLV, the same sub-TLV but with the rightmost bit of the first byte set to one. This sub-TLV is used to acknowledge the completion message. A router MAY send a HELLO PDU immediately in response to a receive completion message or wait for the next scheduled Hello transmission. A router MAY retransmit the same completion message in several HELLO PDUs to ensure that they are correctly received by the neighbour.

5. Security considerations

This document raises no new security issues for IS-IS. For the ordered FIB capability sub-TLV, the discussion in [ISISCAP] is applicable. The authentication mechanisms defined in [RFC] can be used to authenticate the Hello PDUs that carry the completion messages if required.

6. IANA considerations

IANA will assign a new IS-IS sub-TLV code-point for the order FIB capability sub-TLV that can appear inside the Router Capability TLV defined in [ISISCAP].

IANA will assign a new IS-IS TLV code-point for the completion TLV. This TLV can only appear inside Hello PDUs.

Name	Value	IIH	LSP	SNP
Completion messages TLV	TBD	y	n	n

IANA will maintain a registry with the sub-TLVs defined for the completion messages TLV. This document defines three sub-TLVs :

Name	Value
Normal metric change sub-TLV	TBD
Wide metric change sub-TLV	TBD
Overload change sub-TLV	TBD

7. Conclusion

In this document, we have proposed a set of IS-IS TLVs that allows routers using IS-IS to support the ordered FIB convergence and the completion messages proposed in [RTG-OFIB].

References

[IS-IS] "Intermediate system to Intermediate system routeing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)," ISO/IEC 10589:2002, Second Edition.

[IPFRR] M. Shand, S. Bryant, IP Fast Reroute Framework, draft-ietf-rtgwg-ipfrr-framework-04.txt, October 2005

[MPLSFRR] Pan, P. et al, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090.

[PFOB05] P. Francois, O. Bonaventure. Avoiding transient loops during IGP convergence. IEEE INFOCOM 2005, March 2005, Miami, Fl., USA

[RTG-OFIB] P. Francois, O. Bonaventure, M. Shand, S. Previdi, S. Bryant, Loop-free convergence using ordered FIB updates, Internet draft, draft-francois-ordered-fib-00.txt, work in progress, February 2006

[ISISCAP] J.-P. Vasseur, S. Previdi, M. Shand, IS-IS extensions for advertising router capabilities, Internet draft, draft-vasseur-isis-caps-06.txt, February 2004, work in progress

[RFC3567] T. Li, R. Atkinson, Intermediate System to Intermediate System (IS-IS), Cryptographic Authentication, July 2003, RFC3567

Authors Addresses

Olivier Bonaventure, Pierre Francois
Universite catholique de Louvain
Email: FirstName.LastName@uclouvain.be
URL : http://www.info.ucl.ac.be/people/OBO/

Stefano Previdi, Mike Shand
Cisco Systems
Email: {sprevidi,mshand}@cisco.com

Intellectual Property Statement

might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.