Routing Area Working Group
Internet-Draft
Intended status: Informational
Expires: February 19, 2016

P. Sarkar, Ed.
H. Gredler
S. Hegde
C. Bowers
Juniper Networks, Inc.
B. Decraene
Orange

August 18, 2015

LFA selection for Multi-Homed Prefixes
draft-psarkar-rtgwg-multihomed-prefix-lfa-02

Abstract

This document shares experience gained from implementing algorithms to determine Loop-Free Alternates for multi-homed prefixes. In particular, this document provides explicit inequalities that can be used to evaluate neighbors as a potential alternates for multi-homed prefixes. It also provides detailed criteria for evaluating potential alternates for external prefixes advertised by OSPF ASBRs.

Requirements Language

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 RFC2119 [RFC2119].

Status of This Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on February 19, 2016.

Copyright Notice

Copyright (c) 2015 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

The use of Loop-Free Alternates (LFA) for IP Fast Reroute is specified in [RFC5286]. Section 6.1 of [RFC5286] describes a method to determine loop-free alternates for a multi-homed prefixes (MHPs). This document describes a procedure using explicit inequalities that can be used by a computing router to evaluate a neighbor as a potential alternate for a multi-homed prefix. The results obtained are equivalent to those obtained using the method described in Section 6.1 of [RFC5286]. However, some may find this formulation useful.

Section 6.3 of [RFC5286] discusses complications associated with computing LFAs for multi-homed prefixes in OSPF. This document provides detailed criteria for evaluating potential alternates for external prefixes advertised by OSPF ASBRs, as well as explicit inequalities.

2. LFA inequalities for MHPs

This document proposes the following set of LFA inequalities for selecting the most appropriate LFAs for multi-homed prefixes (MHPs). They can be derived from the inequalities in [RFC5286] combined with the observation that D_opt(N,P) = Min (D_opt(N,PO_i) + cost(PO_i,P)) over all PO_i

```
Link-Protection:
D_opt(N,PO_i)+ cost(PO_i,P) < D_opt(N,S) +
                              D_opt(S,PO_best) + cost(PO_best,P)

Link-Protection + Downstream-paths-only:
D_opt(N,PO_i)+ cost(PO_i,P) < D_opt(S,PO_best) + cost(PO_best,P)

Node-Protection:
D_opt(N,PO_i)+ cost(PO_i,P) < D_opt(N,E) +
                              D_opt(E,PO_best) + cost(PO_best,P)

Where,
   S             - The computing router
   N             - The alternate router being evaluated
   E             - The primary next-hop on shortest path from S to
                   prefix P.
   PO_i          - The specific prefix-originating router being
                   evaluated.
   PO_best       - The prefix-originating router on the shortest path
                   from the computing router S to prefix P.
   Cost (X,P)    - Cost of reaching the prefix P from prefix
                  originating node X.
   D_opt(X,Y)    - Distance on the shortest path from node X to node
                   Y.
```

Figure 1: LFA inequalities for MHPs

3. LFA selection for the multi-homed prefixes

To compute a valid LFA for a given multi-homed prefix P, through an alternate neighbor N a computing router S MUST follow one of the appropriate procedures below.

```
Link-Protection :
=================
1. If alternate neighbor N is also prefix-originator of P,
   1.a. Select N as a LFA for prefix P (irrespective of
        the metric advertised by N for the prefix P).
2. Else, evaluate the link-protecting LFA inequality for P with
   the N as the alternate neighbor.
   2.a. If LFA inequality condition is met,
        select N as a LFA for prefix P.
   2.b. Else, N is not a LFA for prefix P.

Link-Protection + Downstream-paths-only :
========================================
1. Evaluate the link-protecting + downstream-only LFA inequality
   for P with the N as the alternate neighbor.
   1.a. If LFA inequality condition is met,
        select N as a LFA for prefix P.
   1.b. Else, N is not a LFA for prefix P.

Node-Protection :
=================
1. If alternate neighbor N is also prefix-originator of P,
   1.a. Select N as a LFA for prefix P (irrespective of
        the metric advertised by N for the prefix P).
2. Else, evaluate the apporpriate node-protecting LFA inequality
   for P with the N as the alternate neighbor.
   2.a. If LFA inequality condition is met,
        select N as a LFA for prefix P.
   2.b. Else, N is not a LFA for prefix P.
```

Figure 2: Rules for selecting LFA for MHPs

In case an alternate neighbor N is also one of the prefix-originators of prefix P, N MAY be selected as a valid LFA for P.

However if N is not a prefix-originator of P, the computing router SHOULD evaluate one of the corresponding LFA inequalities, as mentioned in Figure 1, once for each remote node that originated the prefix. In case the inequality is satisfied by the neighbor N router S MUST choose neighbor N, as one of the valid LFAs for the prefix P.

When computing a dowstream-only LFA, in addition to being a prefix-originator of P, router N MUST also satisfy the downstream-only LFA inequality specified in Figure 1.

For more specific rules please refer to the later sections of this document.

4. LFA selection for the multi-homed external prefixes

Redistribution of external routes into IGP is required in case of two different networks getting merged into one or during protocol migrations. External routes could be distributed into an IGP domain via multiple nodes to avoid a single point of failure.

During LFA calculation, alternate LFA next-hops to reach the best ASBR could be used as LFA for the routes redistributed via that ASBR. When there is no LFA available to the best ASBR, it may be desirable to consider the other ASBRs (referred to as alternate ASBR hereafter) redistributing the external routes for LFA selection as defined in [RFC5286] and leverage the advantage of having multiple re-distributing nodes in the network.

4.1. IS-IS

LFA evaluation for multi-homed external prefixes in IS-IS is similar to the multi-homed internal prefixes. Inequalities described in sec 2 would also apply to multi-homed external prefixes as well.

4.2. OSPF

Loop free Alternates [RFC 5286] describes mechanisms to apply inequalities to find the the loop free alternate neighbor. For the selection of alternate ASBR for LFA consideration, additional rules have to be applied in selecting the alternate ASBR due to the external route calculation rules imposed by [RFC 2328].

This document also defines the inequalities defined in RFC [5286] specifically for the alternate loop-free ASBR evaluation.

4.2.1. Rules to select alternate ASBR

The process to select an alternate ASBR is best explained using the rules below. The below process is applied when primary ASBR for the concerned prefix is chosen and there is an alternate ASBR originating same prefix.

```
   1. If RFC1583Compatibility is disabled

             1a. if primary ASBR and alternate ASBR are intra area
                     non-backbone path go to step 2.
             1b. If primary ASBR and alternate ASBR belong to
                     intra-area backbone and/or inter-area path go
                     to step 2.
             1c. for other paths, skip the alternate ASBR and
                     consider next ASBR.

   2. If cost type (type1/type2) advertised by alternate
      ASBR same as primary
           2a. If  not same skip alternate ASBR and consider next ASBR.

   3. If cost type is type1
             3a. If cost is same, program ECMP
             3b. else go to step 5.

   4  If cost type is type 2
             4a. If cost is different, skip alternate ASBR and
                     consider next ASBR
             4b. If type2 cost is same, compare type 1 cost.
             4c. If type1 cost is also same program ECMP.
             4d. If type 1 cost is different go to step 5.

   5. If route type (type 5/type 7)
              5a. If route type is same, check route p-bit,
                      forwarding address field for routes from both
                      ASBRs
                      match. If not skip alternate ASBR and consider
                      next ASBR.
              5b. If route type is not same, skip ASBR
                      and consider next ASBR.

    6. Apply inequality on the alternate ASBR.
```

Figure 3: Rules for selecting alternate ASBR in OSPF

4.2.2. Multiple ASBRs belonging different area

When "RFC1583compatibility" is set to disabled, OSPF[RFC2328] defines certain rules of preference to choose the ASBRs. While selecting alternate ASBR for loop evaluation for LFA, these rules should be applied and ensured that the alternate neighbor does not loop the traffic back.

When there are multiple ASBRs belonging to different area advertising the same prefix, pruning rules as defined in RFC 2328 section 16.4.1

are applied. The alternate ASBRs pruned using above rules are not considered for LFA evaluation.

4.2.3. Type 1 and Type 2 costs

If there are multiple ASBRs not pruned via rules defined in 3.2.2, the cost type advertised by the ASBRs is compared. ASBRs advertising Type1 costs are preferred and the type2 costs are pruned. If two ASBRs advertise same type2 cost, the alternate ASBRs are considered along with their type1 cost for evaluation.If the two ASBRs with same type2 as well as type1 cost, ECMP FRR is programmed.If there are two ASBRs with different type2 cost, the higher cost ASBR is pruned.The inequalities for evaluating alternate ASBR for type 1 and type 2 costs are same, as the alternate ASBRs with different type2 costs are pruned and the evaluation is based on equal type 2 cost ASBRS.

4.2.4. RFC1583compatibility is set to enabled

When RFC1583Compatibility is set to enabled, multiple ASBRs belonging to different area advertising same prefix are chosen based on cost and hence are valid alternate ASBRs for the LFA evaluation.

4.2.5. Type 7 routes

Type 5 routes always get preference over Type 7 and the alternate ASBRs chosen for LFA calculation should belong to same type.Among Type 7 routes, routes with p-bit and forwarding address set have higher preference than routes without these attributes. Alternate ASBRs selected for LFA comparison should have same p-bit and forwarding address attributes.

4.2.6. Inequalities to be applied for alternate ASBR selection

The alternate ASBRs selected using above mechanism described in 3.2.1, are evaluated for Loop free criteria using below inequalities.

4.2.6.1. Forwarding address set to non zero value

```
Link-Protection:
F_opt(N,PO_i)+ cost(PO_i,P) < D_opt(N,S) +
                              F_opt(S,PO_best) + cost(PO_best,P)

Link-Protection + Downstream-paths-only:
F_opt(N,PO_i)+ cost(PO_i,P) < F_opt(S,PO_best) + cost(PO_best,P)

Node-Protection:
F_opt(N,PO_i)+ cost(PO_i,P) < D_opt(N,E) +
                              F_opt(E,PO_best) + cost(PO_best,P)

Where,
S             - The computing router
N             - The alternate router being evaluated
E             - The primary next-hop on shortest path from S to
                           prefix P.
PO_i          - The specific prefix-originating router being
                           evaluated.
PO_best       - The prefix-originating router on the shortest path
                           from the computing router S to prefix P.
cost(X,Y)     - External cost for Y as advertised by X
F_opt(X,Y)    - Distance on the shortest path from node X to Forwarding
                address specified by ASBR Y.
D_opt(X,Y)    - Distance on the shortest path from node X to node Y.
```

Figure 4: LFA inequality definition when forwarding address in non-zero

4.2.6.2. ASBRs advertising type1 and type2 cost

```
Link-Protection:
D_opt(N,PO_i)+ cost(PO_i,P) < D_opt(N,S) +
                              D_opt(S,PO_best) + cost(PO_best,P)

Link-Protection + Downstream-paths-only:
D_opt(N,PO_i)+ cost(PO_i,P) < D_opt(S,PO_best) + cost(PO_best,P)

Node-Protection:
D_opt(N,PO_i)+ cost(PO_i,P) < D_opt(N,E) +
                              D_opt(E,PO_best) + cost(PO_best,P)

Where,
S             - The computing router
N             - The alternate router being evaluated
E             - The primary next-hop on shortest path from S to
                             prefix P.
PO_i          - The specific prefix-originating router being
                             evaluated.
PO_best       - The prefix-originating router on the shortest path
                             from the computing router S to prefix P.
cost(X,Y)     - External cost for Y as advertised by X.
D_opt(X,Y)    - Distance on the shortest path from node X to node Y.
```

Figure 5: LFA inequality definition for type1 and type 2 cost

5. Acknowledgements

Thanks to Alia Atlas and Salih K A for their useful feedback and inputs.

6. IANA Considerations

N/A. - No protocol changes are proposed in this document.

7. Security Considerations

This document does not introduce any change in any of the protocol specifications. It simply proposes additional inequalities for selecting LFAs for multi-homed prefixes.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

8.2. Informative References

[RFC5286] Atlas, A., Ed. and A. Zinin, Ed., "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, DOI 10.17487/RFC5286, September 2008, <http://www.rfc-editor.org/info/rfc5286>.

Authors' Addresses

Pushpasis Sarkar (editor)
Juniper Networks, Inc.
Electra, Exora Business Park
Bangalore, KA 560103
India

Email: psarkar@juniper.net

Hannes Gredler
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: hannes@juniper.net

Shraddha Hegde
Juniper Networks, Inc.
Electra, Exora Business Park
Bangalore, KA 560103
India

Email: shraddha@juniper.net

Chris Bowers
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: cbowers@juniper.net

Bruno Decraene
Orange

Email: bruno.decraene@orange.com