

Routing Area Working Group  
Internet-Draft  
Intended status: Informational  
Expires: January 9, 2017

P. Sarkar, Ed.  
Individual  
S. Hegde  
C. Bowers  
Juniper Networks, Inc.  
U. Chunduri, Ed.  
Ericsson Inc.  
J. Tantsura  
Individual  
B. Decraene  
Orange  
H. Gredler  
Unaffiliated  
July 8, 2016

LFA selection for Multi-Homed Prefixes  
draft-psarkar-rtwg-multihomed-prefix-lfa-04

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 January 9, 2017.

#### Copyright Notice

Copyright (c) 2016 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. Acronyms	3
2. LFA inequalities for MHPs	4
3. LFA selection for the multi-homed prefixes	4
3.1. Improved coverage with simplified approach to MHPs	6
3.2. IS-IS ATT Bit considerations	7
4. LFA selection for the multi-homed external prefixes	8
4.1. IS-IS	8
4.2. OSPF	8
4.2.1. Rules to select alternate ASBR	8
4.2.2. Multiple ASBRs belonging different area	9
4.2.3. Type 1 and Type 2 costs	10
4.2.4. RFC1583compatibility is set to enabled	10
4.2.5. Type 7 routes	10
4.2.6. Inequalities to be applied for alternate ASBR selection	10
4.2.6.1. Forwarding address set to non zero value	10
4.2.6.2. ASBRs advertising type1 and type2 cost	11
5. LFA Extended Procedures	12
5.1. Links with IGP MAX_METRIC	12
5.2. Multi Topology Considerations	13
6. Acknowledgements	14
7. IANA Considerations	14
8. Security Considerations	14
9. References	14
9.1. Normative References	14
9.2. Informative References	15
Authors' Addresses	15

## 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.

This document also provide clarifications, additional considerations to [RFC5286], to address a few coverage and operational observations. These observations are in the area of handling IS-IS attach (ATT) bit in Level-1 (L1) area, links provisioned with MAX\_METRIC for traffic engineering (TE) purposes and in the area of Multi Topology (MT) IGP deployments. All these are elaborated in detail in Section 3.2 and Section 5.

### 1.1. Acronyms

AF	-	Address Family
ATT	-	IS-IS Attach Bit
ECMP	-	Equal Cost Multi Path
IGP	-	Interior Gateway Protocol
IS-IS	-	Intermediate System to Intermediate System
OSPF	-	Open Shortest Path First
MHP	-	Multi-homed Prefix
MT	-	Multi Topology
SPF	-	Shortest Path First PDU

## 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  $D_{opt}(N,P) = \text{Min} (D_{opt}(N,PO_i) + \text{cost}(PO_i,P))$  over all  $PO_i$

Link-Protection:

$$D_{opt}(N,PO_i) + \text{cost}(PO_i,P) < D_{opt}(N,S) + D_{opt}(S,PO_{best}) + \text{cost}(PO_{best},P)$$

Link-Protection + Downstream-paths-only:

$$D_{opt}(N,PO_i) + \text{cost}(PO_i,P) < D_{opt}(S,PO_{best}) + \text{cost}(PO_{best},P)$$

Node-Protection:

$$D_{opt}(N,PO_i) + \text{cost}(PO_i,P) < D_{opt}(N,E) + D_{opt}(E,PO_{best}) + \text{cost}(PO_{best},P)$$

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.
- $\text{Cost}(X,P)$  - 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 appropriate 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 downstream-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.

3.1. Improved coverage with simplified approach to MHPs

LFA base specification [RFC5286] Section 6.1 recommends that a router compute the alternate next-hop for an IGP multi-homed prefix by considering alternate paths via all routers that have announced that prefix and the same has been elaborated with appropriate inequalities in the above section. However, [RFC5286] Section 6.1 also allows for the router to simplify the multi-homed prefix calculation by assuming that the MHP is solely attached to the router that was its pre-failure optimal point of attachment, at the expense of potentially lower coverage. If an implementation chooses to simplify the multi-homed prefix calculation by assuming that the MHP is solely attached to the router that was its pre-failure optimal point of attachment, the procedure described in this memo can potentially improve coverage for equal cost multi path (ECMP) MHPs without incurring extra computational cost.

While the approach as specified in [RFC5286] Section 6.1 last paragraph, is to simplify the MHP as solely attached to the router that was its pre-failure optimal point of attachment; though it is a scalable approach and simplifies computation, [RFC5286] notes this may result in little less coverage.

This memo improves the above approach to provide loop-free alternatives without any additional cost for equal cost multi path MHPs as described through the below example network. The approach specified here MAY also be applicable for handling default routes as explained in Section 3.2.

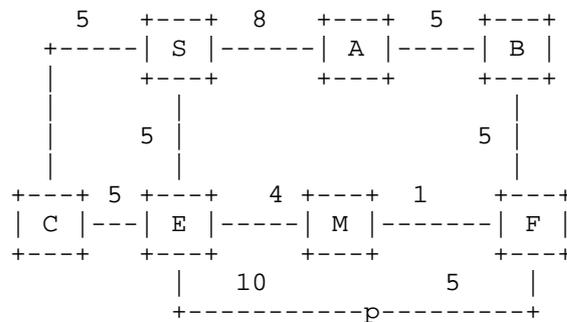


Figure 3: MHP with same ECMP Next-hop

In the above network a prefix p, is advertised from both Node E and Node F. With simplified approach taken as specified in [RFC5286] Section 6.1, prefix p will get only link protection LFA through the neighbor C while a node protection path is available through neighbor

A. In this scenario, E and F both are pre-failure optimal points of attachment and share the same primary next-hop. Hence, an implementation MAY compare the kind of protection A provides to F (link-and-node protection) with the kind of protection C provides to E (link protection) and inherit the better alternative to prefix p and here it is A.

However, in the below network prefix p has an ECMP through both node E and node F with cost 20. Though it has 2 pre-failure optimal points of attachment, the primary next-hop to each pre-failure optimal point of attachment is different. In this case, prefix p shall inherit corresponding LFA to each primary next-hop calculated for the router advertising the same respectively (node E's and node F's LFA).

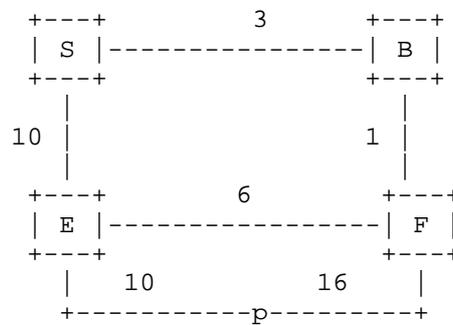


Figure 4: MHP with different ECMP Next-hops

In summary, if there are multiple pre-failure points of attachment for a MHP and primary next-hop of a MHP is same as that of the primary next-hop of the router that was pre-failure optimal point of attachment, an implementation MAY provide the better protection to MHP without incurring any additional computation cost.

### 3.2. IS-IS ATT Bit considerations

Per [RFC1195] a default route needs to be added in Level1 (L1) router to the closest reachable Level1/Level2 (L1/L2) router in the network advertising ATT (attach) bit in its LSP-0 fragment. All L1 routers in the area would do this during the decision process with the next-hop of the default route set to the adjacent router through which the closest L1/L2 router is reachable. The base LFA specification [RFC5286] does not specify any procedure for computing LFA for a default route in IS-IS L1 area. Potentially one MAY consider a default route is being advertised from the border L1/L2 router where ATT bit is set and can do LFA computation for the default route.

But, when multiple ECMP L1/L2 routers are reachable in an L1 area corresponding best LFAs SHOULD be given for each primary next-hop associated with default route. Considerations as specified in Section 3 and Section 3.1 are applicable for default routes, if the default route is considered as ECMP MHP.

#### 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 redistributing 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 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 5: 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:

$$F_{\text{opt}}(N, PO_i) + \text{cost}(PO_i, P) < D_{\text{opt}}(N, S) + F_{\text{opt}}(S, PO_{\text{best}}) + \text{cost}(PO_{\text{best}}, P)$$

Link-Protection + Downstream-paths-only:

$$F_{\text{opt}}(N, PO_i) + \text{cost}(PO_i, P) < F_{\text{opt}}(S, PO_{\text{best}}) + \text{cost}(PO_{\text{best}}, P)$$

Node-Protection:

$$F_{\text{opt}}(N, PO_i) + \text{cost}(PO_i, P) < D_{\text{opt}}(N, E) + F_{\text{opt}}(E, PO_{\text{best}}) + \text{cost}(PO_{\text{best}}, P)$$

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<sub>i</sub> - The specific prefix-originating router being evaluated.
- PO<sub>best</sub> - 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<sub>opt</sub>(X,Y) - Distance on the shortest path from node X to Forwarding address specified by ASBR Y.
- D<sub>opt</sub>(X,Y) - Distance on the shortest path from node X to node Y.

Figure 6: LFA inequality definition when forwarding address in non-zero

#### 4.2.6.2. ASBRs advertising type1 and type2 cost

Link-Protection:

$$D_{\text{opt}}(N, PO_i) + \text{cost}(PO_i, P) < D_{\text{opt}}(N, S) + D_{\text{opt}}(S, PO_{\text{best}}) + \text{cost}(PO_{\text{best}}, P)$$

Link-Protection + Downstream-paths-only:

$$D_{\text{opt}}(N, PO_i) + \text{cost}(PO_i, P) < D_{\text{opt}}(S, PO_{\text{best}}) + \text{cost}(PO_{\text{best}}, P)$$

Node-Protection:

$$D_{\text{opt}}(N, PO_i) + \text{cost}(PO_i, P) < D_{\text{opt}}(N, E) + D_{\text{opt}}(E, PO_{\text{best}}) + \text{cost}(PO_{\text{best}}, P)$$

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<sub>i</sub> - The specific prefix-originating router being evaluated.
- PO<sub>best</sub> - 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<sub>opt</sub>(X,Y) - Distance on the shortest path from node X to node Y.

Figure 7: LFA inequality definition for type1 and type 2 cost

## 5. LFA Extended Procedures

This section explains the additional considerations in various aspects as listed below to the base LFA specification [RFC5286].

### 5.1. Links with IGP MAX\_METRIC

Section 3.5 and 3.6 of [RFC5286] describes procedures for excluding nodes and links from use in alternate paths based on the maximum link metric (as defined in for IS-IS in [RFC5305] or as defined in [RFC3137] for OSPF). If these procedures are strictly followed, there are situations, as described below, where the only potential alternate available which satisfies the basic loop-free condition will not be considered as alternative.



determined by the presence of IGP adjacency from Source to the neighboring node on that MT-ID apart from the administrative restrictions and other checks laid out in [RFC5286]. The same is also applicable for OSPF [RFC4915] [MT-OSPF] or different AFs in multi instance OSPFv3 [RFC5838].

However for MT IS-IS, if a default topology is used with MT-ID 0 [RFC5286] and both IPv4 [RFC5305] and IPv6 routes/AFs [RFC5308] are present, then the condition of network congruency is applicable for LFA computation as well. Network congruency here refers to, having same address families provisioned on all the links and all the nodes of the network with MT-ID 0. Here with single decision process both IPv4 and IPv6 next-hops are computed for all the prefixes in the network and similarly with one LFA computation from all eligible neighbors per [RFC5286], all potential alternatives can be computed.

## 6. Acknowledgements

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

## 7. IANA Considerations

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

## 8. Security Considerations

This document does not introduce any change in any of the protocol specifications and also this does not introduce any new security issues other than as noted in the LFA base specification [RFC5286].

## 9. References

### 9.1. Normative References

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <<http://www.rfc-editor.org/info/rfc1195>>.

[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>>.

## 9.2. Informative References

- [I-D.ietf-rtgwg-lfa-manageability]  
Litkowski, S., Decraene, B., Filsfils, C., Raza, K., and M. Horneffer, "Operational management of Loop Free Alternates", draft-ietf-rtgwg-lfa-manageability-11 (work in progress), June 2015.
- [RFC3137] Retana, A., Nguyen, L., White, R., Zinin, A., and D. McPherson, "OSPF Stub Router Advertisement", RFC 3137, DOI 10.17487/RFC3137, June 2001, <<http://www.rfc-editor.org/info/rfc3137>>.
- [RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, DOI 10.17487/RFC4915, June 2007, <<http://www.rfc-editor.org/info/rfc4915>>.
- [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <<http://www.rfc-editor.org/info/rfc5120>>.
- [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>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<http://www.rfc-editor.org/info/rfc5305>>.
- [RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, DOI 10.17487/RFC5308, October 2008, <<http://www.rfc-editor.org/info/rfc5308>>.
- [RFC5838] Lindem, A., Ed., Mirtorabi, S., Roy, A., Barnes, M., and R. Aggarwal, "Support of Address Families in OSPFv3", RFC 5838, DOI 10.17487/RFC5838, April 2010, <<http://www.rfc-editor.org/info/rfc5838>>.

## Authors' Addresses

Pushpasis Sarkar (editor)  
Individual

Email: [pushpasis.ietf@gmail.com](mailto:pushpasis.ietf@gmail.com)

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

Uma Chunduri (editor)  
Ericsson Inc.  
300 Holger Way,  
San Jose, California 95134  
USA

Phone: 408 750-5678  
Email: uma.chunduri@ericsson.com

Jeff Tantsura  
Individual

Email: jefftant.ietf@gmail.com

Bruno Decraene  
Orange

Email: bruno.decraene@orange.com

Hannes Gredler  
Unaffiliated

Email: hannes@gredler.at