LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 6, 2021

A. Wang
China Telecom
Z. Hu
Y. Xiao
Huawei Technologies
G. Mishra
Verizon Inc.

November 2, 2020

Prefix Unreachable Announcement
draft-wang-lsr-prefix-unreachable-annoucement-04

Abstract

This document describes a mechanism that can be used to announce the unreachable prefixes called PUA (Prefix Unreachable Announcement) in any OSPF multi area or ISIS multi level hierarchy where area summarizations exist.

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 https://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 May 6, 2021.

Table of Contents

1. Introduction

As part of an operators optimized design criteria, a critical requirement is to limit SPF churn which occurs within a single OSPF area or ISIS level. This is accomplished by sub-dividing the IGP domain into multiple areas for flood reduction of intra area prefixes so they are contained within each discrete area to avoid domain wide flooding.

OSPF and ISIS have a default and summary route mechanism which is performed on the OSPF area border router or ISIS L1-L2 node. The OSPF summary route is triggered to be advertise conditionally when at least one component prefix exists within the non-zero area. ISIS Level-L1-L2 node as well generate a summary prefix into the level-2 backbone area for Level 1 area prefixes that is triggered to be advertised conditionally when at least a single component prefix exists within the Level-1 area. ISIS L1-L2 node with attach bit set also generates a default route into each Level-1 area along with summary prefixes generated for other Level-1 areas.

Operators have historically relied on MPLS architecture which is based on LSP FEC binding for service traffic, which is not influenced by the above summary action. But SRV6 routing framework utilities the IPv6 data plane standard IGP LPM (Longest prefix match), when

operators started to migration from MPLS LSP based host route bootstrapped FEC binding, to SRv6 routing framework, the summary action will influence the forwarding of traffic when there is link or node failure within the IGP area.

This document describes a mechanism that can be used to announce the unreachable prefixes called PUA (Prefix Unreachable Announcement) in any OSPF multi area or ISIS multi level hierarchy where area summarizations exist, so service flows can be received at the customer receiver endpoint instead of being dropped at the border node endpoint where the source resides, thereby eliminating an aggregation of excessive drops on the border node.

2. 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 [RFC2119] .

3. Scenario Description

Figure1 illustrates the topology scenario when OSPF is running in multi-area. R0-R4 are routers in backbone area, S1-S4,T1-T4 are internal routers in area 1 and area 2 respectively. R1 and R3 are area border routers between area 0 and area 1. R2 and R4 are area border routers between area 0 and area 2. Ps2 is the host address of S2 and Pt2 is the host address T2.

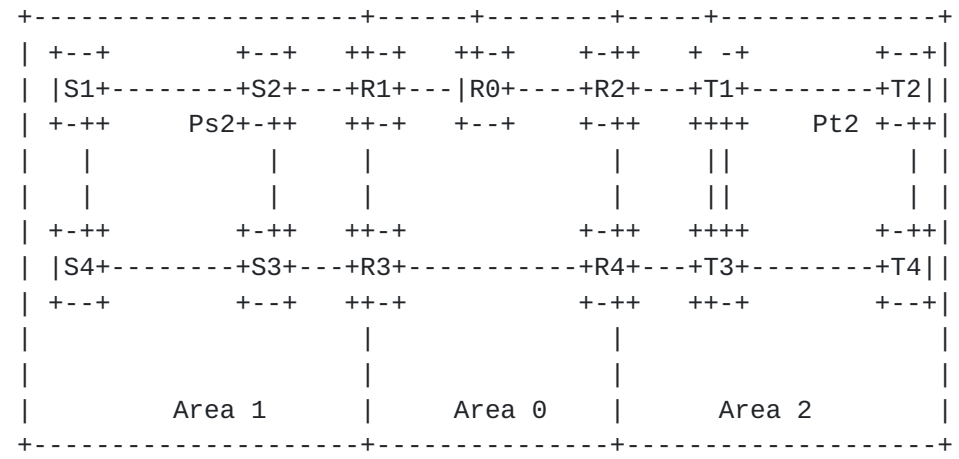


Figure 1: OSPF Inter-Area Prefix Unreachable Announcement Scenario

3.1. Inter-Area Node Failure Scenario

If the area border router R2/R4 does the summary action, then one summary address that cover the prefixes of area 2 will be announced to area 0 and area 1, instead of the detail address. When the node T2 is down, Pt2 become unreachable. But there will be no change to the summary prefix. Except the border router R2/R4, the other routers within area 0 and area 1 do not know the unreachable status of this prefix. When these routers send traffic to prefix Pt2, the traffic will be dropped.

3.2. Inter-Area Links Failure Scenario

In other situation, if the link between T1/T2 and T1/T3 are broken, R2 will not be able to reach node T2. But as R2 and R4 do the summary announcement, and the summary address covers the prefix of Pt2, other nodes in area 0 area 1 will still send traffic to T2 via the border router R2. When R2 receives such traffic, it will drop the packet.

In such situation, the border router R2 should notify other routers that it can't reach the prefix Pt2, and lets the other routers to select R4 as the bypass router to reach prefix Pt2.

3.3. Intra-Area Node Failure Scenario

For intra area, there are some situations that the border routers, for example R1/R3 in Figure 1, announces the summary address that cover also the prefix addresses in area 1. In this situation, when node S2 failures, node S1 should send traffic to the back up path that bypass the failure node. But the back up path can't be triggered because node S1 still think it can reach the prefix Ps2 because it has the summary route that announced by the border router R1/R3.

From the above scenarios, we can conclude that in such situations, the mechanism for Prefix Unreachable Announcement (PUA) should be designed to alleviate the traffic loss.

4. Inter-area prefix unreachable solution

[RFC7794] and [I-D.ietf-lsr-ospf-prefix-originator] both define one sub-TLV "Prefix Source Router ID" to announce the originator router information of one prefix. This TLV can be used to announce the prefix unreachable information when the link or node is down.

According to the procedure described in section 5 of [I-D.ietf-lsr-ospf-prefix-originator], the ABR has the responsibility

to add the prefix originator information when it receive the Router LSA from other routers in the same area. When the ABR does the summary work and receives one updated LSA that omits the prefix belong to failed link which is within the range of summary address, the ABR should announce one new Summary LSA, which includes the information about this prefix, but with the prefix originator set to NULL(all 0 address).

When one node in one area is down, the ABR has also the ability to detect the missing neighbor from the neighbor list. It should then announce one new Summary LSA that includes the loopback addresses of this node, with the prefix originator set also to NULL(all 0 address).

For IS-IS, the above procedure is similar. The level-1/2 router will accomplish the above work when it judges that one prefix within the summary address range is missing.

These LSAs will be transported via the traditional flooding procedure.

When the routers in other area receives such LSA, they will generate automatically one black-hole route, with the prefix as the destination, and the next hop be set to Null. If there is other router advertise the summary prefix without carry unreachable information, it will prefer the other router to reach the specified prefix.

5. Intra-area prefix unreachable solution

In the intra-area scenario, like S1 illustrated in Figure 1, it will learn two types of prefixes, one is summary route, another is host route. When node S2 is down, S2 will withdraw the host route. But S1 can still match the summary route via the longest mask matching. For this scenario, when node S2 is down, S1 needs to keep the S2 host route for a period of time but updates S2 host route to black hole route. S1 will match the black hole route via the longest mask matching. Such mechanism can be used to trigger a SRv6 VPN for PE switching, or SRv6 TE mid-point protection.

The period for keeping the black hole route should be configured, to ensure the related protocols or services be converged.

6. Implementation Consideration

The above procedures will only be triggered under the following conditions:

1. The ABR or Level 1/2 router do the summary work.

2. The link prefix or loopback address of the node within the summary address range become unreachable.

The Summary LSA that includes the unreachable prefix, with the prefix originator set to NULL value, will be announced across the ABR router, reach the routers in other areas. It's behavior is still the same as that defined in OSPFv2 [RFC2328] or OSPFv3 [RFC5340]

Considering the balances of reachable information and unreachable information announcement capabilities, the implementation of this mechanism should set one MAX_Address_Announcement (MAA) threshold value that can be configurable. Then, the ABR should make the following decisions to announce the prefixes:

1. If the number of unreachable prefixes is less than MAA, the ABR should advertise the summary address and the PUA.

2. If the number of reachable address is less than MAA, the ABR should advertise the detail reachable address only.

3. If the number of reachable prefixes and unreachable prefixes exceed MAA, then advertise the summary address with MAX metric.

When the receiver receives such LSA, it will do the following judgements and actions:

1. If all the source that announces the summary route announces the prefix unreachable information, the receiver should add the black hole route to this prefix.

2. If not, the receiver should prefer the router that does not include the prefix unreachable information to reach this prefix.

3. The receiver router should keep the black hole routes for PUA as one configurable time(MAX_T_PUA) to allow the services that depends on them converged. After the MAX_T_PUA time, such black hole routes can be deleted then.

6.1. Usages of Tunnel among ABRs

When in situation that all the ABRs reach the announcement limit, it is not viable to increase the cost of summary address, as described in above paragraph. In such situation, the operator should provide other solution to decrease the packet loss that due to the advertised summary route, which includes the failure prefix. Figure 2 illustrate such situation.

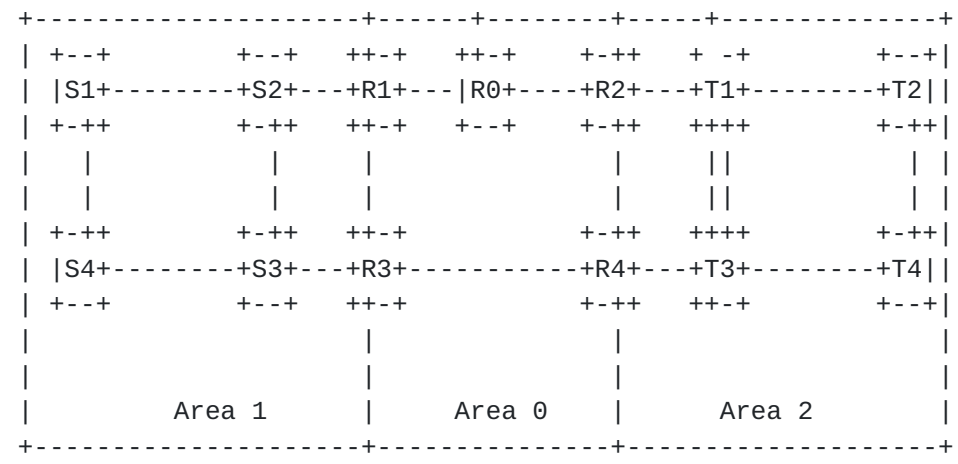


Figure 2: Usage of Tunnel among ABRs

In Figure 2, when R1 and R3 reach the PUA MAA state simultaneously, it is no use for these two ABRs increase the summary cost. For example, when the link between S1 and S4 is down, R1 can reach S1/S2 but not S3/S4, R3 can reach S3/S4 but not S1/S2. If the traffic destined to S3/S4 be sent via R1, it will be dropped by R1, but such traffic can be sent to the destination via R3. The traffic destined to S1/S2 that be sent via R3 will have the same fate.

In such situation, it is useful for R1 to send these traffic via some tunnel to R3 and vice versa. To achieve this, the ABR (R1/R3) should build the tunnel in advance. When one of the ABRs receive the failure information, it should check whether the missed nodes can be reached via other ABRs. If such missed nodes can be reached, it then install the tunnel route as the next hop to these missed nodes. And when it receives the related traffic, it can transfer the traffic via other ABRs. Such implementation can mitigate the traffic loss in Figure 2.

In order to prevent the traffic loop, when one ABR receives such traffic via the tunnel interface but can't find the next hop for these traffic, it should drop such traffic and can't send again via tunnel to other ABRs.

If ABR receive the link/node failure information, and can't find other ABRs to reach the missed nodes, it should send some notify messages to the operator because some nodes are out of the network and the ABRs can't notify the nodes in other area via the PUA mechanism.

6.2. Fast Rerouting to Avoid Routing Backhole

Fast rerouting is a mechanism that allows a router whose local link has failed to forward traffic to a pre-computed alternate path until the router installs the new primary next-hops based upon the changed network topology. If the area border router R2/R4 does the summary action, both R2 and R4 should pre-install one path to the summary address, with the nexthop address pointed to each other. When the ABR R2 becomes unreachable to a node in one area, R2 will withdraw the detailed route of the node. The pre-install summary route will be the longest match route for the summary address. The traffic destined to the failed node arrived on R2 will be forwarded to another ABR R4 then. If R4 have the detailed route of the node, R4 will forward the traffic to the corresponding node along the correct path. When both R2 and R4 becomes unreachable to the failed node, how to avoid the traffic loops between R2 and R4 is beyond the scope of this document. .

6.3. PUA Capabilities Announcement

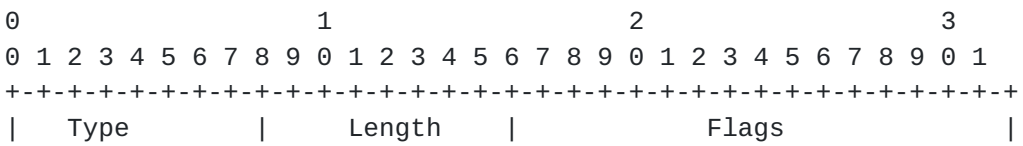
When not all of the nodes in one area support the PUA information, there are possibilities to form traffic loop. To avoid this happen, the ABR should not send PUA information to one area until it ensures that all of nodes in this area can parse the PUA information. In the situation that not all of nodes support PUA information, the ABR should use the mechanism that described section Section 6.1 and Section 6.2 to forward the received traffic that bound to the unreachable prefixes.

To accomplish this, this draft defines the capabilities sub-TLV as the followings:

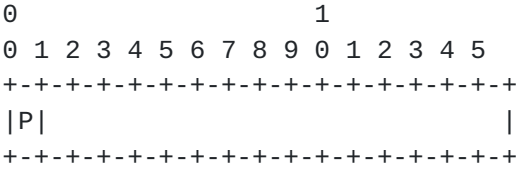
For OSPFv2, this bit (Bit number TBD, suggest bit 6, 0x20) should be carried in "OSPF Router-LSA Option", as that described in RFC2328 [RFC2328]

For OSPFv3, one bit (Bit number TBD, suggest bit 8) should be defined to indicate the router's capabilities to support PUA that described in this draft, the defined bit should be carried in "OSPF Router Informational Capabilities" TLV, which is described in [RFC7770]

For ISIS, one new sub-TLV(Type TBD, suggest 29), PUA Capabilities sub-TLV, which is included in the "IS-IS Router CAPABILITY TLV" [RFC7981] is defined in the followings:



Type: TBD, Suggested value 29, to be assigned by IANA
Length: 2
Flags: 2 octets
The following flags are defined:



where:

P-flag: If set, the router supports PUA information.

Figure 3: PUA Capabilities sub-TLV format

7. Security Considerations

Security concerns for OSPF are addressed in [RFC5709]

Advertisement of the additional information defined in this document may raise some compatible issues when the node does not recognize it or consider such information is illegal. During deployment, the operator should make sure all the routers within its domain have supported such features.

8. IANA Considerations

IANA is requested to register the following in the "OSPF Router Properties Registry" and "OSPF Router Informational Capability Bits Registry" respectively.

Bit Number	Capability Name	Reference
TBD(0x20)	OSPF PUA Support	this document

Table 1: P-Bit in OSPF Router-LSA Option

Bit Number	Capability Name	Reference
TBD(bit 8)	OSPF PUA Support	this document

Table 2: OSPF Router PUA Capability Support Bit

IANA is requested to register the following in "Sub-TLVs for TLV242(IS-IS Router CAPABILITY TLV)

Type: 29 (Suggested - to be assigned by IANA)

Description: PUA Support Capabilities

9. Acknowledgement

Thanks Peter Psenak, Les Ginsberg and Acee Lindem for their suggestions and comments on this draft.

10. Normative References

[I-D.ietf-lsr-ospf-prefix-originator]
Wang, A., Lindem, A., Dong, J., Psenak, P., and K. Talaulikar, "OSPF Prefix Originator Extensions", draft-ietf-lsr-ospf-prefix-originator-07 (work in progress), October 2020.

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

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <https://www.rfc-editor.org/info/rfc2328>.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <https://www.rfc-editor.org/info/rfc5340>.

[RFC5709] Bhatia, M., Manral, V., Fanto, M., White, R., Barnes, M., Li, T., and R. Atkinson, "OSPFv2 HMAC-SHA Cryptographic Authentication", RFC 5709, DOI 10.17487/RFC5709, October 2009, <https://www.rfc-editor.org/info/rfc5709>.

[RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 7770, DOI 10.17487/RFC7770, February 2016, <https://www.rfc-editor.org/info/rfc7770>.

[RFC7794] Ginsberg, L., Ed., Decraene, B., Previdi, S., Xu, X., and U. Chunduri, "IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability", RFC 7794, DOI 10.17487/RFC7794, March 2016, <https://www.rfc-editor.org/info/rfc7794>.

[RFC7981] Ginsberg, L., Previdi, S., and M. Chen, "IS-IS Extensions for Advertising Router Information", RFC 7981, DOI 10.17487/RFC7981, October 2016, <https://www.rfc-editor.org/info/rfc7981>.

Authors' Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing 102209
China

Email: wangaj3@chinatelecom.cn

Zhibo Hu
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: huzhibo@huawei.com

Yaqun Xiao
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: xiaoyaqun@huawei.com

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring MD 20904
United States of America

Email: gyan.s.mishra@verizon.com