LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: 22 December 2023

A. Wang
China Telecom
Z. Hu
Huawei Technologies
J. Sun
ZTE Corporation
C. Lin
New H3C Technologies

20 June 2023

Prefix Unreachable Announcement
draft-wang-lsr-prefix-unreachable-annoucement-12

Abstract

This document describes a mechanism that can trigger the switchover of the services which rely on the reachability of the peer endpoints, for example the BGP or the tunnel services. It is mainly used in the scenarios that the summary prefixes are advertised at the border routers whereas the services endpoints are located in different IGP areas or levels, whose reachabilities are covered by the summary prefixes.

It introduces a new signaling mechanism using a negative prefix announcement called Prefix Unreachable Announcement Mechanism(PUAM), utilized to detect a link or node down event and signal the overlay services that the communication endpoint is unreachabe to force the overlay service headend switchover immediately.

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 22 December 2023.

Table of Contents

1. Introduction

As part of an operator optimized design, a critical requirement is to limit Shortest Path First (SPF) churn which occurs within a single OSPF area or IS-IS 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 IS-IS have a default and summary route mechanism which is performed on the OSPF area border router or IS-IS L1-L2 node. The summary route is triggered to be advertised conditionally when at least one component prefix exists within the attached area or Level.

Operators have historically relied on MPLS architecture which is based on exact match host route for single area. LDP inter-area extension [RFC5283] provides the ability to LPM(Longest Prefix Match), so now it can be a summary match of a host route of the egress PE for an inter-area LSP to be instantiated.

SRV6 routing framework utilities the IPv6 data plane standard IGP LPM, such summarization will influence the forwarding of traffic when a link or node failure event occurs for a component prefix within the summary range, result in black hole routing of traffic.

The motivation behind this draft is for either MPLS LPM FEC binding, SRv6 etc. tunnel ,or BGP overlay service that are using LPM forwarding plane where the IGP domain has been carved up into OSPF areas or IS-IS levels and summarization is utilized. In such scenario, a link or node failure can result in a black hole of traffic when the summary advertisement that covers the failure prefixes still exists.

PUAM can track the unreachabilities of the important component prefixes to ensure traffic is not black hole sink routed for the above overlay services.

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

Figure 1 illustrates the topology scenario when OSPF or IS-IS is running in multi areas. 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.

S1/S4 and T2/T4 PEs peer to customer CEs for overlay VPNs. Ps1/Ps4 is the loopback0 address of S1/S4 and Pt2/Pt4 is the loopback0 address of T2/T4.

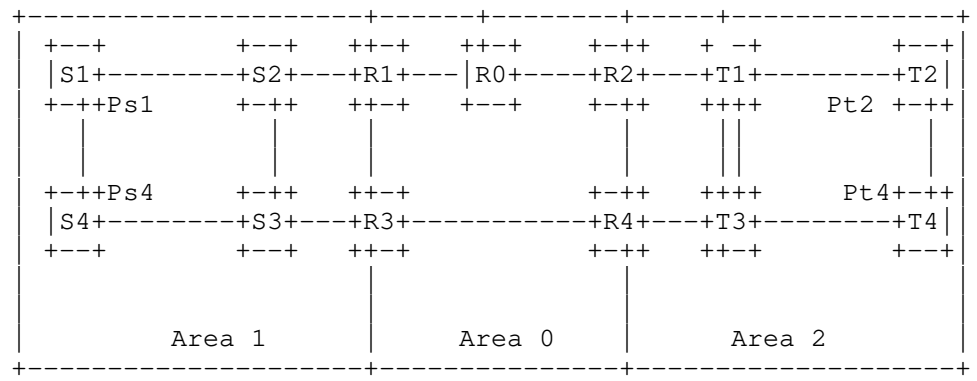


Figure 1: OSPF Inter-Area Prefix Unreachable Announcement Scenario

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 becomes unreachable while the summary prefix continues to be advertised into the backbone area. Except the border router R2/R4, the other routers within area 0 and area 1 do not know the unreachable status of the Pt2 prefix. Traffic will continue to forward via LPM match to prefix Pt2 and will be dropped on the ABR node, resulting in black hole routing and connectivity loss. Even the customer overlay VPN are dual homed to both S1/S4 and T2/R4, traffic will not be able to fail-over to alternate egress PE(T4) due to the summarization.

3.2. Inter-Area Links Failure Scenario

In a link failure scenario, if the links between T1/T2 and T1/T3 are down, 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 and area 1 will still send traffic to T2 via the border router R2, thus black hole sink routing the traffic.

In such a situation, the border router R2 should notify other routers that it can’t reach the prefix Pt2, and lets the other ABRs(R4) being selected as the next hop to reach prefix Pt2.

4. PUAM (Prefix Unreachable Advertisement Mechanism) Procedures

[RFC7794] and [RFC9084] define sub-TLV to announce the originator information of the one prefix from a specified node. This draft utilizes such sub-TLV for OSPF and IS-IS to signal the negative prefix in the perspective PUAM when a link or node goes down.

When OSPF ABR or IS-IS L1-L2 border node detects link or node down, the ABR should announce one new summary LSA or LSP which includes the information about the down prefix, with the prefix originator sub-TLV set to NULL(0.0.0.0). The LSA or LSP will be propagated with standard flooding procedures.

If the nodes in the area receive the PUAM message for one prefix from all of its ABR routers, they will know that the specified prefix is unreachable and start overlay services switchover process if such services rely on unreachable prefix. Without the PUAM forced switchover, the summary prefix will yield black hole routing and results in loss of connectivity.

When only some of the ABRs can’t reach the failure node/link, as that described in Section 3.2, along with the PUAM message for the associated prefixe from these ABRs, the ABR that can reach the PUAM prefix should advertise the specific route to this prefix. The internal routers within another area can then bypass the ABRs that can’t reach the PUAM prefix, to reach the prefix that advertised in PUAM message.

5. PUAM Capabilities Announcement

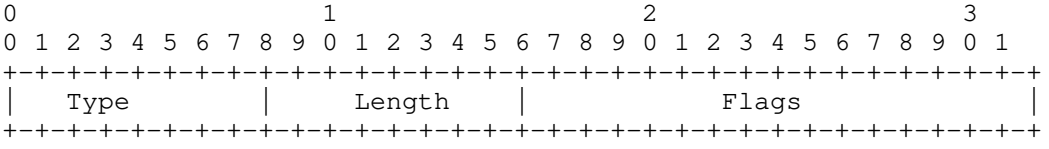
When not all of the nodes in one area support the PUAM information, there are possibilities the nodes misbehavior when they don’t support the received PUAM message.

To avoid this happen, the ABR should know the capabilities of each node within one area via the newly defined capabilities bits, and advertise PUAM message with some additional information when necessary.

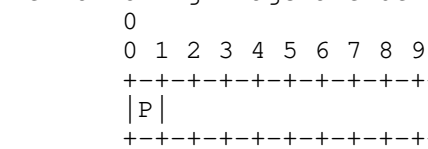
For OSPFv2, this bit (Bit number TBD, suggest bit 6, 0x20) should be carried in "OSPF Router-LSA Option", as that described in [RFC2328].

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

For IS-IS, one new sub-TLV(Type TBD, suggest 29), PUAM 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 PUAM information.

Figure 2: PUAM Capabilities sub-TLV format

If not all of nodes within one area support the PUAM capabilities, the PUAM message should be advertised with the associated prefix cost set to LSInfinity. According to the description in [RFC2328], [RFC5305] and [RFC5308], the prefix advertised with such metric value will not be considered during the normal SPF computation, then such additional information will avoid the misbehavior of the nodes when they don't recognize the PUAM message.

If all of nodes within one area support the PUAM capabilites, the PUAM message can be safely advertised without the additional LSInfinity metric information.

6. Implementation Consideration

Considering the balances of reachable information and unreachable information announcements, 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 PUAM.

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 MAA unreachable prefixes, and also the summary address with MAX(LSInfiity-1) metric. At the same time, the ABR should notify the operators there are severe incident occurs within the network.

7. Deployment Considerations

To support the PUAM advertisement, the ABRs should be upgraded according to the procedures described in Section 4. The nodes that want to accomplish the services switchover should also be upgraded to act upon the receive of the PUAM message. Other nodes within the network can ignore such PUAM message if they don't care or don't support it.

As described in Section 4, the ABR will advertise the PUAM message once it detects there is link or node down within the summary address. In order to reduce the unnecessary advertisements of PUAM messages on ABRs, the ABRs should support the configuration of the tracked prefixes. Based on such information, the ABR will only advertise the PUAM message when the tracked prefixes(for example, the loopback addresses of PEs that run BGP) that within the summary address is missing.

The advertisement of PUAM message should only last one configurable period to allow the services that run on the failure prefixes are switchovered.

If one prefix is missed before the PUAM takes effect, the ABR will not declare its absence via the PUAM.

8. Security Considerations

Advertisement of PUAM information follow the same procedure of traditional LSA. The action based on the PUAM is depended on the overlay services and is out of the scope of this document.

There is no changes to the forward behavior of other internal routers.

9. 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 PUAM Support	this document

Table 1: P-Bit in OSPFv2 Router-LSA Option

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

Table 2: OSPFv3 Router PUAM 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: PUAM Support Capabilities

10. Acknowledgement

Thanks Peter Psenak, Les Ginsberg, Bruno Decraene, Acee Lindem, Shraddha Hegde, Robert Raszuk, Tony Li, Jeff Tantsura and Tony Przygienda for their suggestions and comments on this draft.

11. 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, <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>.

[RFC5283] Decraene, B., Le Roux, JL., and I. Minei, "LDP Extension for Inter-Area Label Switched Paths (LSPs)", RFC 5283, DOI 10.17487/RFC5283, July 2008, <https://www.rfc-editor.org/info/rfc5283>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <https://www.rfc-editor.org/info/rfc5305>.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, DOI 10.17487/RFC5308, October 2008, <https://www.rfc-editor.org/info/rfc5308>.

[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>.

[RFC9084] Wang, A., Lindem, A., Dong, J., Psenak, P., and K. Talaulikar, Ed., "OSPF Prefix Originator Extensions", RFC 9084, DOI 10.17487/RFC9084, August 2021, <https://www.rfc-editor.org/info/rfc9084>.

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

Jinsong
ZTE Corporation
No. 68, Ziijnhua Road
Nanjing
210012
China
Email: sun.jinsong@zte.com.cn

Changwang
New H3C Technologies
Beijing
China
Email: linchangwang.04414@h3c.com