L   P. Shaofu
Internet-Draft   C. Ran
Intended status: Standards Track   ZTE Corporation
Expires: March 9, 2021   G. Mirsky
                                                              ZTE Corp.
                                                      September 5, 2020

IGP Flexible Algorithm Optimazition for Netwrok Slicing
draft-peng-lsr-flex-algo-opt-slicing-02

Abstract

IGP Flex Algorithm proposes a solution that allows IGPs themselves to compute constraint based paths over the network, and it also specifies a way of using Segment Routing (SR) Prefix-SIDs and SRv6 locators to steer packets along the constraint-based paths. This document extends the use of the IGP Flex Algorithm to to be more suitable for network slicing scenarios.

Table of Contents

1. Introduction

IGP Flex Algorithm [I-D.ietf-lsr-flex-algo] proposes a solution that allows IGPs themselves to compute constraint based paths over the network, and it also specifies a way of using Segment Routing (SR) Prefix-SIDs and SRv6 locators to steer packets along the constraint-based paths. It specifies a set of extensions to ISIS, OSPFv2 and OSPFv3 that enable a router to send TLVs that identify (a) calculation-type, (b) specify a metric-type, and (c)describe a set of constraints on the topology, that are to be used to compute the best paths along the constrained topology. A given combination of calculation-type, metric-type, and constraints is known as an FAD (Flexible Algorithm Definition).

[I-D.peng-teas-network-slicing] proposes a solution to extend the control plane of transport network to instantiate the Network Slice Instance (NSI) in transport network. A new identifier, AII, instead of existing TE affinity or other identifiers, is introduced to represent a TN-slice and specify the dedicated resource for the TN-slice.

This document extends the FAD of IGP Flex Algorithm to let IGPs compute constraint based paths limited in specific TN-slice.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. SR Policy Using Slice-based Resources

[I-D.ietf-spring-segment-routing-policy] details the concepts of SR Policy and steering into an SR Policy. These apply equally to the MPLS and IPv6 (known as SRv6) data plane instantiations of Segment Routing with their respective representations of segments as SR-MPLS SID and SRv6 SID as described in [RFC8402]. The color of SR policy defines a TE purpose, which includes a set of constraints such as bandwidth, delay, TE metric, etc.

The overlay service can select underlay SR policy according to a meaningful color value. From the perspective of service, color is the key to get the expected SLA, and it is a global administrative configuration or setting that could be exchangeable between two devices for SR policy on-demand next-hop triggering. The service never concern whether the underlay network has been partitioned as multi-domains, or multi-topologies. That is, color has not semantic local within one domain, or one topology. Instead, any type of resources such as topology, computation, storage could be selected by the color template. In this sense, TN-slices are also high-level resouces that could be selected by color template. A simple way to achieve this is to contain the specific AII information in the color template, to restrict the TE path to the corresponding TN-slice.

4. SR Policy Optimaztion with IGP Flex-algo

Indeed, FA-id defined in [I-D.ietf-lsr-flex-algo] is a short mapping of SR policy color to optimaze segment stack depth for the IGP area partial of the entire SR policy. The overlay service that want to be carried over a particual SR-FA path must firstly let the SR policy supplier know that requirement. There are two possible ways to map a color to an FA-id. One is explicit mapping configuration within color template, the other is dynamic to replace a long segment list to a single FA segment by headend or controller once the creteria contained in the color-template equal to that contained in FAD.

In addtion to the above mapping mode, merging mode is also possible. In this case, the computation engine will combine the constraints contained in the expanded FAD with other constraints in the color template. This is to continue to iterate TE business in FA plane. The iteration means the best-effort path itself within an FA plane is exactly constraint based path, but operators can define more constraints in that FA plane. The computed strict path can be optimized to a loose path when a part of the strict path is consistent with the algorithm based path, i.e, some consecutive adjacency SIDs can be replaced with a single algorithm based prefix SID. Note that the loose optimization in this case, i.e, an SR policy created in FA plane, is similar with that when an SR policy is created in physical network, and that is different with optimization of segment stack depth using Flex-algo.

[I-D.ietf-lsr-flex-algo] described that application specific Flex-Algorithm participation advertisements MAY be topology specific or MAY be topology independent, and also emphasize that for Segment Routing application, the Flex-Algorithm participation advertisement is topology independent, i.e., when a router advertises participation in an SR-Algorithm, the participation applies to all topologies in which the advertising node participates. Here the topology means Multi-Topology Routing (MTR) described in [RFC5120], [RFC4915], [RFC5340]. [RFC8402] also mentioned that multiple SIDs MAY be allocated to the same prefix so long as the tuple <prefix, topology, algorithm> is unique. In fact, this will lead to many forwarding tables, such as table per MTR, table per each combined tuple <MTR, algorithm>, and make traffic steering very complicated.

According to [I-D.peng-teas-network-slicing], we donot use MTR to identify the TN-slice and partition the virtual topology for the TN-slice. Instead, a slice-based identifier, AII, is introduced to represent a TN-slice. The first feature of AII is a TE criteria for TE service just like AG/EAG. So that AII, like other constraints, can be included in color template. When an SR policy uses Flex-algo for stack depth optimization, in order to make the contents of the color template and the mapping FAD consistent, AII is also necessary put into the mapping FAD.

Although the network operator may change the AII information within the FAD for the specific FA-id, there is only one forwarding table with constant table ID, i.e., FA-id. Note that there are also independent forwarding tables per AII for other purpose, but not those per tuple <AII, FA-id>. That is, FA-id has not semantic local within AII, just as color is not part of the topology.

5. IGP Flex-algo Enhancement with AII

FAD that contains AII information will enhance the capability of Flex-algo to support network slicing.

5.1. Enhancement for TI-LFA

Loop Free Alternate (LFA) paths for a given Flex-Algorithm can include Prefix-SIDs advertised specifically for the given algorithm, and especially Adjacency-SIDs for the specific AII. When different FA planes share the same link resouce, Adjacency-SID per AII (according to [I-D.peng-teas-network-slicing]) can distinguish the flow of different slices well and provide different treatment.

The following figure shows an example of Flex-algo enhancement with AII.

```
[S1]--------[D]--------[S2]
 |           |          |
 |           |          |
 |           |          |
[A]---------[B]--------[C]
```

Figure 1: Flex-algo LFA Enhancement with AII

Suppose that node S1, A, B, D and their inter-connected links belongs to FA-id 128 plane as well as AII-1, and S2, B, C, D and their inter-connected links belongs to FA-id 129 plane as well as AII-2. The IGP metric of link B-D is 100, and all other links have IGP metric 1. In FA-id 128 plane, from S1 to destination D, the primary path is S1-D, and the TI-LFA backup path is segment list {node(B), adjacency(B-D)}. Similarly, In FA-id 129 plane, from S2 to destination D, the primary path is S2-D, and the TI-LFA backup path is segment list {node(B), adjacency(B-D)}. With the help of AII parameter contained in the FAD, the above TI-LFA path of FA-id 128 plane will be translated to {node-SID(B)@FA-id128, adjacency-SID(B-D)@AII-1}, and TI-LFA path of FA-id 129 plane will be translate to {node-SID(B)@FA-id129, adjacency-SID(B-D)@AII-2}. So that node B can distinguish the flow of FA-id 128 and FA-id 129 with different treatment (e.g., QoS) and send to the same outgoing link B-D.

5.2. Enhancement for Inter-domain

For inter-domain case, different domain can config different FA-id independently, but they can contain the same AII to construct an E2E slice-based SR policy. IGP flex-algo is responsible for creating constraint based paths within the domain according to FAD including

AII parameter, and BGP-LU or SDN controller is responsible for selecting inter-domain links according to color template including AII parameter. AII is easy to address the requirement of E2E Slicing view.

5.3. Enhancement for L2bundles

[RFC8668] inroduced L2 Bundle Member Attributes sub-TLV to be advertised within ISIS TLV-22/23/141/222/223. It may define an attribute common to all of the bundle members listed and an attribute individual to each L2 Bundle Member. [RFC8668] mainly defined Adjacency-SID for each L2 Bundle Member, that is very useful to isolate flows among different slices. A typical deployment of hard slicing is that different L2 Bundle Member, e.g, Flex-E channel, belongs to different TN-slice, so a specific Adjacency-SID for a specific L2 Bundle Member will steer the packets to that member.

However, the link resource of an FA plane is selected according to the TE affinity attribute of all L3 links joining to the IGP instance and the INCLUDE/EXCLUDE rules contained in the FAD. If we want to let different L2 Bundle Member belongs to different FA plane, it must determine the TE affinity of each L2 Bundle Member. There could have two methods:

- o [I-D.zch-lsr-isis-network-slicing] extends ISIS protocol to carry AII (TN-slice ID) information for each L2 Bundle Member within L2 Bundle Member Attributes sub-TLV defined in [RFC8668]. Even though the same L3 parrent link can be joined to multiple FA planes according to TE affinity INCLUDE/EXCLUDE rules, it is easy for each FA plane to exactly use the specific L2 Bundle Member according to the AII information contained in FAD.

- o It is also possible to directly define AG/EGA per L2 Bundle Member for Flex-algo application. [I-D.peng-lsr-flex-algo-l2bundles] described how to create Flex-algo plane in the case of L2bundles scenario, by explicit AG/EGA affinity configuration for each L2 Bundle Member. This method is similar with the above definition of AII per L2 Bundle Member, both them can distinguish traffic of different hard slice.

It is possible to include only AII, or only TE affinity, or both AII and TE affinity, in an FAD.

6. QoS Policy per AII/Algorithm

Due to SID allocation per algorithm, flows belonging to different FA planes can be easily distinguished by incoming SID of the received

packets, so that different QoS policies can be applied to different FA packets on the same link.

Depending on the implementation, operators can configure multiple QoS policies each for different algorithm on the same link. One of the difficulties is that during this configuration phase it is not straightforward for a link to be included in an FA plane, as this can only be determined after all nodes in the network have negotiated the FAD. A simple way is that as long as a node enable an FA, all its links are configured with that algorithm based QoS policy.

Depending on the implementation, operators can also configure multiple QoS policies each for different AII on the same link. An AII based QoS policy is configured to a subset of links who join the related TN-slice explicitly.

Depending on the implementation, the queue resources of the link can be divided according to AII, algorithm, or their combination. When the received pachet matched an flex-algo related FIB entry, it will be directed to the queue dedicated to that algorithm. If the FIB entry is created according to the FAD including AII creteria, the packets can further schedule queue resouce according to AII.

7. AII of FAD Sub-TLV

7.1. ISIS AII of FAD Sub-TLV

ISIS AII of FAD Sub-TLV is used to advertise the AII information that is used during the Flex-Algorithm path calculation. It is a Sub-TLV of the ISIS FAD Sub-TLV. It has the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              AII                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: ISIS AII of FAD Sub-TLV format

where:

Type: TBD1.

Length: 4 octets.

AII: Administrative Instance Identifier as defined in [I-D.peng-teas-network-slicing].

ISIS AII of FAD Sub-TLV MAY NOT appear more then once in an ISIS FAD Sub-TLV. If it appears more then once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

7.2. OSPF AII of FAD Sub-TLV

OSPF AII of FAD Sub-TLV is used to advertise the AII information that is used during the Flex-Algorithm path calculation. It is a Sub-TLV of the OSPF FAD TLV. It has the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Type             |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              AII                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: OSPF AII of FAD Sub-TLV format

where:

Type: TBD2.

Length: 4 octets.

AII: Administrative Instance Identifier as defined in [I-D.peng-teas-network-slicing].

OSPF AII of FAD Sub-TLV MAY NOT appear more then once in an OSPF FAD TLV. If it appears more then once, the OSPF FAD TLV MUST be ignored by the receiver.

8. IANA Considerations

8.1. ISIS IANA Considerations

This document defines the following Sub-Sub-TLVs in the "Sub-Sub-TLVs for Flexible Algorithm Definition Sub-TLV" registry:

Type: TBD1

Description: Administrative Instance Identifier

Reference: This document (Section 6.1)

8.2. OSPF IANA Considerations

This document registers following Sub-TLVs in the "TLVs for Flexible Algorithm Definition TLV" registry:

Type: TBD2

Description: Administrative Instance Identifier

Reference: This document (Section 6.2)

9. Security Considerations

This specification inherits all security considerations of [I-D.ietf-lsr-flex-algo].

10. Acknowledgements

TBD

11. Normative References

[I-D.ietf-lsr-flex-algo]
Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", draft-ietf-lsr-flex-algo-10 (work in progress), August 2020.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-08 (work in progress), July 2020.

[I-D.peng-lsr-flex-algo-l2bundles]
Peng, S., Chen, R., and G. Mirsky, "IGP Flexible Algorithm with L2bundles", draft-peng-lsr-flex-algo-l2bundles-02 (work in progress), August 2020.

[I-D.peng-teas-network-slicing]
Peng, S., Chen, R., Mirsky, G., and F. Qin, "Packet Network Slicing using Segment Routing", draft-peng-teas-network-slicing-03 (work in progress), February 2020.

[I-D.zch-lsr-isis-network-slicing]
Zhu, Y., Chen, R., Peng, S., and F. Qin, "IS-IS Extensions to Support Transport Network Slices using Segment Routing", draft-zch-lsr-isis-network-slicing-06 (work in progress), September 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>.

[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, <https://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, <https://www.rfc-editor.org/info/rfc5120>.

[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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8668] Ginsberg, L., Ed., Bashandy, A., Filsfils, C., Nanduri, M., and E. Aries, "Advertising Layer 2 Bundle Member Link Attributes in IS-IS", RFC 8668, DOI 10.17487/RFC8668, December 2019, <https://www.rfc-editor.org/info/rfc8668>.

Authors' Addresses

Peng Shaofu
ZTE Corporation
No.68 Zijinghua Road, Yuhuatai District
Nanjing
China

Email: peng.shaofu@zte.com.cn

Chen Ran
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing
China

Email: chen.ran@zte.com.cn

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com