Networking Working Group Shaofu. Peng
Internet-Draft Ran. Chen
Intended status: Standards Track Gregory. Mirsky
Expires: August 29, 2019 ZTE Corporation
February 25, 2019

Packet Network Slicing using Segment Routing
draft-peng-lsr-network-slicing-00

Abstract

This document presents a mechanism aimed at providing a solution for network slicing in the transport network for 5G services. The proposed mechanism uses a unified administrative instance identifier to distinguish different virtual network resources for both intra-domain and inter-domain network slicing scenarios. Combined with the segment routing technology, the mechanism could be used for both best-effort and traffic engineered services for tenants.

Table of Contents

1. Introduction

According to 5G context, network slicing is the collection of a set of technologies to create specialized, dedicated logical networks as a service (NaaS) in support of network service differentiation and meeting the diversified requirements from vertical industries. Through the flexible and customized design of functions, isolation mechanisms, and operation and management (O&M) tools, network slicing is capable of providing dedicated virtual networks over a shared infrastructure. A Network slice instance (NSI) is the realization of network slicing concept. It is an E2E logical network, which comprises of a group of network functions, resources, and connection relationships. An NSI typically covers multiple technical domains, which includes a terminal, access network (AN), transport network (TN) and a core network (CN), as well as DC domain that hosts third-party applications from vertical industries. Different NSIs may have different network functions and resources. They may also share some of the network functions and resources.

For a packet network, network slicing requires the underlying network to support partitioning of the network resources to provide the

client with dedicated (private) networking, computing, and storage resources drawn from a shared pool. The slices may be seen as virtual networks. [I-D.ietf-teas-enhanced-vpn] described a framework to create virtual networks in a packet network. This document specifies a detailed mechanism to signal association of shared resources required to create and manage an NSI.

Currently there are multiple methods that could be used to identify the virtual network resource, such as Administrative Group (AG) described in [RFC3630], [RFC5329] and [RFC5305], Extended Administrative Groups (EAGs) described in [RFC7308], and Multi-Topology Routing (MTR) described in [RFC5120], [RFC4915] and [RFC5340]. However, all these methods are not sufficient to meet the requirements of network slicing service. For example, AG or EAG are limited to serve as a link color scheme used in TE path computation to meet the requirements of TE service for a tenant. But it is difficult to use them for an NSI allocation mapping (assuming that each bit position of AG/EAG represents an NSI) and, at the same time, as an IGP FIB identifier for best-effort service for the same tenant. MTR is limited to serve as an IGP logical topology scheme only used in the intra-domain scenario, and it is challenging to select inter-area link resource according to MT-ID when E2E inter-domain TE path needs to be created for a tenant.

Thus, there needs to be a new characteristic of NSI to isolate underlay resources. Firstly it could serve as TE criteria for TE service, and secondly, as an IGP FIB table identifier for best-effort service. This document introduces a new property of NSI called "Administrative Instance Identifier" (AII) and corresponding method of using it.

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. Overview of Mechanism

At the initial stage, each link in a physical network can be colored to conform with network slicing requirements. As previously mentioned, AII can be used to color links to partition underlay resource. Also, we may continue to use AG or EAG to color links for traditional TE purpose within a virtual network specified by an AII. A single or multiple AIIs could be configured on each intra-domain or inter-domain link regardless of IGP instance configuration. At the minimum, a link always belongs to default AII (the value is 0). The number of AIIs configured on a node's links determines the number of

virtual networks the node is part of. This document defines a new extension of the existing IGP-TE mechanisms [RFC3630] and [RFC5305] to distribute AII information in an AS as a new TE parameter of a link. An SDN controller, using BGP-LS or another interface, will have a distinct view of each virtual network specified by AII.

Using the CSPF algorithm, a TE path for any best-effort (BE) or traffic engineered (TE) service can be calculated within a virtual network specified by the AII. The computation criteria could be <AII, min igp-metric> or <AII, traditional TE critierias> for the BE and TE respectively. Combined with segment routing, the TE path could be represented as:

- o a single node-SID of the destination node, for the best-effort service in the domain;
- o node-SIDs of the border node and the destination node, adjacency-SID of inter-domain link, for the inter-domain best-effort service;
- o an adjacency-SID list, for P2P traffic engineered service.

Because packets of the best-effort service could be transported over an MP2P LSP without congestion control, SR best-effort FIB for each virtual network specified by AII to forward best-effort packets may be created in the IGP domain. Thus, CSPF computation with criteria <AII, min igp-metric>is distributed on each node in the IGP domain. That is similar to the behavior in [Flex-algo], but the distributed CSPF computation is triggered by AII.

To distinguish forwarding behavior of different virtual networks, prefix-SID need to be allocated per AII and advertised in the IGP domain.

For inter-domain case, in addition to the destination node-SID, several node-SIDs of the domain border node and adjacency-SID of inter-domain link are also needed to construct the E2E segment list. The segment list could be computed with the help of the SDN controller which needs to consider AII information during the computation. The head-end of the segment list maintains the corresponding SR-TE tunnel or [I-D.ietf-spring-segment-routing-policy].

As for the prefix-SID, adjacency-SID needs to be allocated per AII to distinguish the forwarding behavior of different virtual networks.

For P2P traffic engineering service, especially such as uRLLC service, it SHOULD not transfer over an MP2P LSP to avoid the risk of

traffic congestion. The segment list could consist of pure adjacency-SID per AII specific. The head-end of the segment list maintains the corresponding SR-TE tunnel or [I-D.ietf-spring-segment-routing-policy].

However, label stack depth of the segment list MAY be optimized at a later time based on local policies.

At this moment we can steer traffic of overlay service to the above SR best-effort FIB, SR-TE tunnel or SR policy instance, for the specific virtual network. The overlay service could specify a color for TE purpose, for example, color 1000 means <AII=10, min igp-metric> to say that I need best-effort forwarding within AII 10 resource, color 1001 means <AII=10, delay=10ms, AG=0x1> to say that I need traffic engineering forwarding within AII 10 resource, especially using link with AG equal to 0x1 to reach guarantee of not exceeding 10ms delay time. Service with color 1000 will be steered to an SR best-effort FIB entry, or an SR-TE tunnel/policy in case of inter-domain. Service with color 1001 will be steered to an SR-TE tunnel/policy.

4. Resource Allocation per AII

4.1. L3 Link Resource AII Configuration

In IGP domain, each numbered or unnumbered L3 link could be configured with AII information and synchronized among IGP neighbors. The IGP link-state database will contain L3 links with AII information to support TE path computation considering AII criteria. For a numbered L3 link, it could be represented as a tuple <local node-id, remote node-id, local ip-address, remote ip-address>, for unnumbered it could be <local node-id, remote node-id, local interface-id, remote interface-id>. Each L3 link could be configured to belong to a single AII or multiple AIIs, for each <L3 link, AII> tuple it would allocate a different adjacency-SID. Note that an L3 link always belongs to default AII(0).

An L3 link that is not part of the IGP domain, such as the special purpose for a static route, or an inter-domain link, can also be configured with AII information and allocate adjacency-SID per AII as the same as IGP links. BGP-LS could be used to collect link state data with AII information to the controller.

4.2. L2 Link Resource AII Configuration

[I-D.ietf-isis-l2bundles] described how to encode adjacency-SID for each L2 member link of an L3 parent link. It is beneficial to deploy LAG or another virtual aggregation interface in network slicing

scenario. Between two nodes, the dedicated link resources belong to different virtual networks could be added or removed on demand, they are treated as L2 member links of a single L3 virtual interface. It is the single L3 virtual interface that needs to occupy IP resource and be part of the IGP instance. Creating a new slice-specific link on demand or removing the old one, is likely to affect some configurations.

Each L2 member link of an L3 parent link SUGGESTED to be configured to belong to a single AII, and different L2 member link will have different single AII configuration, with different adjacency-SID. Note that in this case, the L3 parent link belongs to default AII(0), but each L2 member link belongs to the specific non-default AII. Routing protocol control packets follow the L3 parent link of the L2 member link with the highest priority. At the same time, data packets that belong to the specific virtual network will pass along the L2 member link with the specific AII value.

TE path computation based on link-state database need view the detailed L2 members of an L3 adjacency to select the expected L2 link resource.

4.3. Node Resource AII Configuration

For topology resource, each node needs to allocate node-SID per AII when it joins the related virtual network. All nodes in the IGP domain can run CSPF algorithm with criteria <AII, min IGP metric> to compute best-effort next-hop to any other destination nodes for a virtual network AII-specific, based on the link-state database that containing AII information so that SR best-effort FIB can be constructed for each AII.

An intra-domain overlay best-effort service belongs to a virtual network could directly match in the above SR best-effort FIB for the specific AII, while an inter-domain overlay best-effort service belongs to a virtual network could be over a segment list containing domain border node-SID and destination node-SID which could match in the above SR best-effort FIB for the specific AII.

5. Interworking with SR Flex-algorithm

[I-D.ietf-lsr-flex-algo] introduced a mechanism to do label stack depth optimization for an SR policy in IGP domain part. As the color of SR policy defined a TE purpose, traditionally the headend or SDN controller will compute an expected TE path to meet that purpose. It is necessary to map a color (32 bits) to an FA-ID (8 bits) when SR flex-algorithm enabled for an SR policy, besides that, it is necessary to enable the FA-ID on each node that will join the same FA

plane manually. However, the FAD could copy the TE constraints contained in the color template. We need to consider the cost of losing the flexibility of color when executing the flex-algo optimization, and also consider the gap between P2P TE requirements and MP2P SR LSP capability, to reach the right balance when deciding which SR policy need optimization.

5.1. Best-effort Service AII-specific

As described above, for best-effort service we have already constructed SR best-effort FIB per AII, that is mostly like [Flex-algo]. Thus, it is not necessary to map to FA-ID again for a color template which has defined a best-effort behavior within the dedicated AII. Of course, if someone forced to remap it, there is no downside for the operation, the overlay best-effort service (with a color which defined specific AII, best-effort requirement, and mapping FA-ID) in IGP domain will try to recurse over <AII, prefix> or <FA-ID, prefix> FIB entry.

5.2. Traffic Engineering service AII-specific

An SR-TE tunnel/policy that served for traffic engineering service of a virtual network specified by an AII was generated and computed according to the relevant color template, which contained specific AII and some other traditional TE constraints. If we config mapping FA-ID under the color template, the SR-TE tunnel/policy instance could inherit forwarding information from corresponding SR Flex-Algo FIB entry.

6. Examples

In this section, we will further illustrate the point through some examples. All examples share the same figure below.


```
               .-----.                                      .-----.
```

```
           (       )                                    (       )
       .--(         )--.                            .--(         )--.
   +---+----link A1----+-----+            +---  +----link A1----+---+
   |PE1|----link A2----|ASBR1|---link A1---|ASBR2|----link A2----|PE2|
   |   |----link B1----|     |---link B1---|     |----link B1----|   |
   +---+----link B2----+-----+            +--  -+----link B2----+---+
     (                   )                    (                   )
       '--(  AS1    )--'                        '--(   AS2    )--'
           (       )                                (       )
            '-----'                                  '-----'
```

Figure 1 Network Slicing via AII

Suppose that each link belongs to separate virtual network, e.g., link Ax belongs to the virtual network colored by AII A, link Bx belongs to the virtual network colored by AII B. link x1 has an IGP metric smaller than link x2, but TE metric lager.

To simplify the use case, each AS just contained a single IGP area.

6.1. intra-domain network slicing

From the perspective of node PE1 in AS1, it will calculate best-effort forwarding entry for each AII instance (including default AII) to destinations in the same IGP area. For example:

For <AII=0, destination=ASBR1> entry, forwarding information could be ECMP during link A1 and link B1, with destination node-SID 100 for <AII=0, destination=ASBR1>.

For <AII=A, destination=ASBR1> entry, forwarding information could be link A1, with destination node-SID 200 for <AII=A, destination=ASBR1>.

For <AII=B, destination=ASBR1> entry, forwarding information could be link B1, with destination node-SID 300 for <AII=B, destination=ASBR1>.

It could also initiate an SR-TE instance (SR tunnel or SR policy) with the particular color template on PE1, PE1 is headend and ASBR1 is destination node. For example:

For SR-TE instance 1 with color template which defined criteria including {default AII, min TE metric}, forwarding information could be ECMP during two segment list {adjacency-SID 1002 for <AII=0, link A2>@PE1} and {adjacency-SID 1004 for <AII=0, link B2>@PE1}.

For SR-TE instance 2 with the color template which defined criteria including {AII=A, min TE metric}, forwarding information could be presented as the segment list {adjacency-SID 2002 for <AII=A, link A2>@PE1}.

For SR-TE instance 3 with the color template which defined criteria including {AII=B, min TE metric}, forwarding information could be presented as the segment list {adjacency-SID 3004 for <AII=B, link B2>@PE1}.

Furthermore, we can use SR Flex-algo to optimize the above SR-TE instance. For example, for SR-TE instance 1, we can define FA-ID 201 with FAD that contains the same information as the color template, in turn, FA-ID 202 for SR-TE instance 2, FA-ID 203 for SR-TE instance 3. Note that each FA-ID also needs to be enabled on ASBR1. So that the corresponding SR FA entry could be:

For <FA-ID=201, destination=ASBR1> entry, forwarding information could be ECMP during link A2 and link B2, with destination node-SID 600 for <FA-ID=201, destination=ASBR1>.

For <FA-ID=202, destination=ASBR1> entry, forwarding information could be link A2, with destination node-SID 700 for <FA-ID=202, destination=ASBR1>.

For <FA-ID=203, destination=ASBR1> entry, forwarding information could be link B2, with destination node-SID 800 for <FA-ID=203, destination=ASBR1>.

6.2. inter-domain network slicing via BGP-LS

An E2E inner-AS SR-TE instance with particular color template could be initiated on PE1, PE1 is head-end and PE2 is destination node. BGP-LS could be used to inform the SDN controller about the underlay network topology information including AII attribute. Thus the controller could calculate E2E TE path within the particular virtual network. For best-effort service, for example:

For SR-TE instance 4 with color template which defined criteria including {default AII, min IGP metric}, forwarding information could be segment list {node-SID 100 for <AII=0, destination=ASBR1>, adjacency-SID 1001 for <AII=0, link A1>@ASBR1, node-SID 400 for < AII=0, destination=PE2>}.

For SR-TE instance 5 with color template which defined criteria including {AII=A, min IGP metric}, forwarding information could be segment list {node-SID 200 for <AII=A, destination=ASBR1>, adjacency-

SID 1001 for <AII=A, link A1>@ASBR1, node-SID 500 for <AII=A, destination=PE2>}.

For SR-TE instance 6 with color template which defined criteria including {AII=B, min IGP metric}, forwarding information could be segment list {node-SID 300 for <AII=B, destination=ASBR1>, adjacency-SID 1003 for <AII=B, link B1>@ASBR1, node-SID 600 for <AII=B, destination=PE2>}.

For TE service, for example:

For SR-TE instance 7 with color template which defined criteria including {default AII, min TE metric}, forwarding information could be ECMP during two segment list {adjacency-SID 1002 for <AII=0, link A2>@PE1, adjacency-SID 1001 for <AII=0, link A1>@ASBR1, adjacency-SID 1002 for <AII=0, link A2>@ASBR2} and {adjacency-SID 1004 for <AII=0, link B2>@PE1, adjacency-SID 1003 for <AII=0, link B1>@ASBR1, adjacency-SID 1004 for <AII=0, link B2>@ASBR2}.

For SR-TE instance 8 with color template which defined criteria including {AII=A, min TE metric}, forwarding information could be segment list {adjacency-SID 2002 for <AII=A, link A2>@PE1, adjacency-SID 2001 for <AII=A, link A1>@ASBR1, adjacency-SID 2002 for <AII=A, link A2>@ASBR2}.

For SR-TE instance 9 with color template which defined criteria including {AII=B, min TE metric}, forwarding information could be segment list {adjacency-SID 3004 for <AII=B, link B2>@PE1, adjacency-SID 3003 for <AII=B, link B1>@ASBR1, adjacency-SID 3004 for <AII=B, link B2>@ASBR2}.

For TE service, if we use SR Flex-algo to do optimizaztion, the above forwarding information of each TE instance could inherit the corresponding SR FA entry, it would look like this:

For SR-TE instance 7, forwarding information could be ECMP during two segment list {node-SID 600 for <FA-ID=201, destination=ASBR1>, adjacency-SID 1001 for <AII=0, link A1>@ASBR1, node-SID 600 for < FA-ID=201, destination=PE2>} and {adjacency-SID 1004 for <AII=0, link

B2>@PE1, adjacency-SID 1003 for <AII=0, link B1>@ASBR1, adjacency-SID 1004 for <AII=0, link B2>@ASBR2}.

For SR-TE instance 8 with color template which defined criteria including {AII=A, min TE metric}, forwarding information could be segment list {node-SID 700 for <FA-ID=202, destination=ASBR1>, adjacency-SID 2001 for <AII=A, link A1>@ASBR1, node-SID 700 for <FA-ID=202, destination=PE2>}.

For SR-TE instance 9 with color template which defined criteria including {AII=B, min TE metric}, forwarding information could be segment list {node-SID 800 for <FA-ID=203, destination=ASBR1>, adjacency-SID 3003 for <AII=B, link B1>@ASBR1, node-SID 800 for <FA-ID=203, destination=PE2>}.

6.3. inter-domain network slicing via BGP-LU

In some deployments, operators adopt BGP-LU to build inter-domain MPLS LSP, overlay service will be directly over BGP-LU LSP. If overlay service has TE requirements that defined by a color, that means that BGP-LU LSP needs to have a sense of color too, i.e., BGP-LU label could be allocated per color. BGP-LU LSP generated for specific color will be over intra-domain SR-TE or SR Best-effort path generated for that color again.

In figure 1, PE2 can allocate and advertise six labels for its loopback plus color 1, 2, 3, 4, 5, 6 respectively. Suppose color 1 defines {default AII, min IGP metric}, color 2 defines {AII=A, min IGP metric}, color 3 defines {AII=B, min IGP metric}, and color 4 defines {default AII, min TE metric}, color 5 defines {AII=A, min TE metric}, color 6 defines {AII=B, min TE metric}. PE2 will advertise these labels to ASBR2 and ASBR2 then continues to allocate six labels each for prefix PE2 plus different color. Other nodes will have the same operation. Ultimately PE1 will maintain six BGP-LU LSP.

For example, the BGP-LU LSP for color 1 will be over SR best-effort FIB entry node-SID 100 for <AII=0, destination=ASBR1> to pass through AS1, over adjacency-SID 1001 for <AII=0, link A1>@ASBR1 to pass inter-AS, over SR best-effort FIB entry node-SID 400 for <AII=0, destination=PE2> to pass through AS2.

For example, The BGP-LU LSP for color 4 will over SR-TE instance 1 (see section 6.1), or SR best-effort FIB entry node-SID 600 for <FA-ID=201, destination=ASBR1> (see section 6.1) to pass through 6AS1, over adjacency-SID 1001 for <AII=0, link A1>@ASBR1 to pass inter-AS, over SR-TE instance 1' or corresponding SR FA entry to pass through AS2.

7. Implementation suggestions

As a node often contains control plane and forwarding plane, a suggestion is that only default AII specific FTN entry need be installed on forwarding plane, so that there are not any modification and upgrade requirement for hardware and existing MPLS forwarding mechanism. FTN entry for non-default AII instance will only be maintained on the control plane and be used for overlay service iteration according to next-hop plus color (color will give AII

information and mapping FA-ID information). Note ILM entry for all AII need be installed on forwarding plane.

The same suggestion is also appropriate for SR Flex-algo.

8. IANA Considerations

TBD.

9. Security Considerations

TBD.

10. Acknowledgements

TBD.

11. Normative references

[I-D.ietf-isis-l2bundles]
Ginsberg, L., Bashandy, A., Filsfils, C., Nanduri, M., and E. Aries, "Advertising L2 Bundle Member Link Attributes in IS-IS", draft-ietf-isis-l2bundles-07 (work in progress), May 2017.

[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-01 (work in progress), November 2018.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d., bogdanov@google.com, b., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-02 (work in progress), October 2018.

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Networks (VPN+) Service", draft-ietf-teas-enhanced-vpn-01 (work in progress), February 2019.

[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>.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <https://www.rfc-editor.org/info/rfc3630>.

[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>.

[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>.

[RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, Ed., "Traffic Engineering Extensions to OSPF Version 3", RFC 5329, DOI 10.17487/RFC5329, September 2008, <https://www.rfc-editor.org/info/rfc5329>.

[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>.

[RFC7308] Osborne, E., "Extended Administrative Groups in MPLS Traffic Engineering (MPLS-TE)", RFC 7308, DOI 10.17487/RFC7308, July 2014, <https://www.rfc-editor.org/info/rfc7308>.

Authors' Addresses

Shaofu Peng
ZTE Corporation

Email: peng.shaofu@zte.com.cn

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.cn

Gregory Mirsky
ZTE Corporation

Email: gregimirsky@gmail.com