SPRING Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 31, 2021

J. Dong
Huawei Technologies
S. Bryant
Futurewei Technologies
T. Miyasaka
KDDI Corporation
Y. Zhu
China Telecom
F. Qin
Z. Li
China Mobile
F. Clad
Cisco Systems

July 30, 2020

Introducing Resource Awareness to SR Segments
draft-ietf-spring-resource-aware-segments-00

Abstract

This document describes the mechanism to associate network resource attributes to Segment Routing Identifiers (SIDs). Such SIDs are referred to as resource-aware SIDs in this document. The resource-aware SIDs retain their original forwarding semantics, but with the additional semantics to identify the set of network resources available for the packet processing action. The resource-aware SIDs can therefore be used to build SR paths or virtual networks with a set of reserved network resources. The proposed mechanism is applicable to both segment routing with MPLS data plane (SR-MPLS) and segment routing with IPv6 data plane (SRv6).

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 RFC 2119 [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 https://datatracker.ietf.org/drafts/current/.

Table of Contents

1. Introduction

Segment Routing (SR) [RFC8402] specifies a mechanism to steer packets through an ordered list of segments. A segment is referred to by its Segment Identifier (SID). With SR, explicit source routing can be achieved without introducing per-path state into the network. Compared with RSVP-TE [RFC3209], currently SR does not have the capability of reserving network resources or identifying a set of network resources reserved for individual services or customers. Although a centralized controller can have a global view of network

state and can provision different services using different SR paths, in data packet forwarding it still relies on traditional DiffServ QoS mechanism [RFC2474] [RFC2475] to provide coarse-grained traffic differentiation in the network. While such kind of mechanism may be sufficient for some types of services, some customers or services may require a set of dedicated network resources to be allocated in the network to achieve resource isolation from other customers/services in the same network. Also note the number of such customers or services can be larger than the number of traffic classes available with DiffServ QoS.

This document extends the SR paradigm without the need of defining new SID types by associating SIDs with network resource attributes. These resource-aware SIDs retain their original functionality, with the additional semantics of identifying the set of network resources available for the packet processing action. On a particular network segment, multiple resource-aware SIDs can be allocated, each of which represents a subset of network resources allocated to meet the requirement of individual customers or services. The allocation of network resources on a network segment can be done via a controller or via local configuration, then each set of resource is associated with a resource-aware SID. These resource-aware SIDs can be used to build SR paths with a set of reserved network resources, which can be used in network scenarios which require to allocate a set of network resources for the processing of groups of service traffic. The resource-aware SIDs can also be used to build SR based virtual networks with the required network topology and resource attributes. The proposed mechanism is applicable to SR with both MPLS data plane (SR-MPLS) and IPv6 data plane (SRv6).

2. Segments with Resource Awareness

In segment routing architecture [RFC8402], several types of segments are defined to represent either topological or service instructions. A topological segment can be a node segment or an adjacency segment. A service segment may be associated with specific service functions for service chaining purposes. This document introduces additional resource semantics to these existing types of SIDs, so that the SIDs can be used to identify the topology or service functions, and the set of network resources allocated on the network segments for packet processing.

This section describes the mechanisms of using SR SIDs to identify the additional resource information of SR paths or virtual networks with the two SR data plane instantiations: SR-MPLS and SRv6. The mechanisms to identify the forwarding path or network topology with a SID as defined in [RFC8402] are unchanged, and the control plane can be based on [RFC4915], [RFC5120] and [I-D.ietf-lsr-flex-algo].

2.1. SR-MPLS

As specified in [RFC8402], an IGP Adjacency Segment (Adj-SID) is an IGP-segment attached to a unidirectional adjacency or a set of unidirectional adjacencies. An IGP Prefix segment is an IGP segment representing an IGP prefix, and IGP node segment is an IGP-Prefix segment that identifies a specific router (e.g., a loopback). As described in [I-D.ietf-spring-segment-routing-central-epe] and [I-D.ietf-idr-bgpls-segment-routing-epe], BGP PeerAdj SID is used as an instruction to steer over a specific local interface towards a specific peer node in a peering Autonomous System (AS). These types of SIDs can be extended to represent both topological elements and the resources allocated on a network segment. The MPLS instantiation of Segment Routing is specified in [RFC8660].

For one IGP link, multiple Adj-SIDs SHOULD be allocated, each of which is associated with a network topology the link participates, and MAY represent a subset of link resources. Several approaches can be used to partition the link resource, such as [FLEXE], Layer-2 logical sub-interfaces, dedicated queues, etc. The detailed mechanism of resource partitioning is out of scope of this document.

Similarly, for one IGP node, multiple prefix-SIDs SHOULD be allocated, each of which is associated with a network topology the node participates, and MAY represent a subset of the node resource (e.g. the processing resources). For one inter-domain link, multiple BGP PeerAdj SIDs SHOULD be allocated, each of which is associated with a specific network topology, which spans multiple domains, and MAY represent a subset of link resource allocated on the inter-domain link. Note that this per-segment resource allocation complies to the SR paradigm, which avoids introducing per-path state into the network.

A group of resource-aware SIDs associated with the same network topology can be used to construct the SR paths (either strict or loose) to steer traffic within the topology. Each SID in the SID-list of the SR path MAY represent the set of network resources reserved on the corresponding network segment.

In data packet forwarding, the SIDs are used to identify the topology the packet belongs to, so that a topology specific next-hop can be determined. In addition, the adj-SIDs MAY also be used to steer traffic of different services into different set of link resources. The prefix-SIDs MAY be used to steer traffic of different services into different set of node resources. When a prefix-SID is used in the SID-list to build an SR loose path, the transit nodes can use the prefix-SID to identify the network topology and the associated group of resource, and can process the packet using the local resources

allocated to the corresponding resource group. Note in this case, it is RECOMMENDED that Penultimate Hop Popping (PHP) [RFC3031] be disabled, otherwise the inner service label SHOULD be used to infer the set of resources to be used on the egress node of the SR path.

This mechanism requires to allocate additional prefix-SIDs or adj-SIDs for network segments to identify different set of network resources. As the number of resource groups increases, the number of SIDs would increase accordingly, while it should be noted that there is no per-path state introduced into the network.

2.2. SRv6

As specified in [I-D.ietf-spring-srv6-network-programming], an SRv6 Segment Identifier (SID) is a 128-bit value which consists of a locator (LOC) and a function (FUNCT), optionally it may also contain additional arguments (ARG) immediately after the FUNCT. The LOC of the SID is routable and leads to the node which instantiates that SID, which means the LOC can be parsed by all nodes in the network. The FUNCT part of the SID is an opaque identification of a local function bound to the SID, which means the FUNCT and ARG parts can only be parsed by the node which instantiates that SID.

Taking the above into consideration, for a network node, multiple SRv6 LOCs SHOULD be allocated, each of which is associated with a network topology, and MAY represent a subset of the network resources associated with a virtual network. The SRv6 SIDs of a particular virtual network SHOULD be allocated from the SID space using the resource-aware LOC as the prefix. These SRv6 SIDs can be used to represent virtual network specific local functions.

A group of SRv6 SIDs associated with the same network topology can be used to construct the SR paths (either strict or loose) to steer the traffic of particular service within the topology. Each SID in the SID-list of the SR path MAY also represent the set of network resources on the corresponding network segment.

In data packet forwarding, the LOC part of SRv6 SID is used by transit nodes to identify the topology the packet belongs to, so that a topology specific next-hop can be determined. The LOC MAY also be used to indicate the set of local network resources on the transit nodes to be used for the forwarding of the received packet. The SRv6 segment endpoint nodes use the SRv6 SID to identify the topology the packet belongs to, and the particular local function to perform on the received packet. The local SRv6 SID MAY also be used to identify the set of network resource to be used for executing the local function.

This mechanism requires to allocate additional SRv6 Locators and SIDs for network segments to identify different set of network resources. As the number of resource groups increases, the number of Locators and SIDs would increase accordingly, while it should be noted that there is no per-path state introduced into the network.

3. Control Plane Considerations

The mechanism described in this document makes use of a centralized controller to collect the information about the network (configuration, state, routing databases, etc.) as well as the service information (traffic matrix, performance statistics, etc.) for the planning of network resources based on service requirement. The controller is also responsible for the centralized computation and optimization of the SR paths with both the topology and network resource constraints. The resource-aware SIDs can be either explicitly provisioned by the controller, or dynamically allocated by network nodes then reported to the controller. The interaction between the controller and the network nodes can be based on PCEP [RFC5440], Netconf/YANG [RFC6241] [RFC7950] and BGP-LS [RFC7752]. In some scenarios, extensions to some of these protocols is needed, which are out of the scope of this document and will be specified in separate documents. In some cases, a centralized controller may not be used, but this would complicate the operations and planning therefore not suggested.

The distributed control plane is complementary to the centralized controller. A distributed control plane can be used for the collection and distribution of the network topology and resource information associated with SIDs among network nodes. Distributed route computation for services with topology and resource constraints may also be needed. The distributed control plane may be based on [RFC4915], [RFC5120], [I-D.ietf-lsr-flex-algo] or the combination of some of them with necessary extensions. The details are out of the scope of this document.

4. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

5. Security Considerations

The security considerations of segment routing are applicable to this document.

The Resource-aware SIDs may be used for provisioning of SR paths or virtual networks to carry traffic with latency as one of the SLA parameters. By disrupting the latency of such traffic an attack can be directly targeted at the customer application, or can be targeted at the network operator by causing them to violate their SLA, triggering commercial consequences. Dynamic attacks of this sort are not something that networks have traditionally guarded against, and networking techniques need to be developed to defend against this type of attack. By rigorously policing ingress traffic and carefully provisioning the resources provided to such services, this type of attack can be prevented. However care needs to be taken when providing shared resources, and when the network needs to be reconfigured as part of ongoing maintenance or in response to a failure.

The details of the underlay network MUST NOT be exposed to third parties, to prevent attacks aimed at exploiting a shared resource.

6. Contributors

Zhenbin Li
Email: lizhenbin@huawei.com

Zhibo Hu
Email: huzhibo@huawei.com

7. Acknowledgements

The authors would like to thank Mach Chen, Stefano Previdi, Charlie Perkins, Bruno Decraene, Loa Andersson, Alexander Vainshtein and Joel Halpern for the valuable discussion and suggestions to this document.

8. References

8.1. 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>.

[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>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

8.2. Informative References

[FLEXE] "Flex Ethernet Implementation Agreement", March 2016, <http://www.oiforum.com/wp-content/uploads/OIF-FLEXE-01.0.pdf>.

[I-D.ietf-idr-bgpls-segment-routing-epe] Previdi, S., Talaulikar, K., Filsfils, C., Patel, K., Ray, S., and J. Dong, "BGP-LS extensions for Segment Routing BGP Egress Peer Engineering", draft-ietf-idr-bgpls-segment-routing-epe-19 (work in progress), May 2019.

[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-08 (work in progress), July 2020.

[I-D.ietf-spring-segment-routing-central-epe] Filsfils, C., Previdi, S., Dawra, G., Aries, E., and D. Afanasiev, "Segment Routing Centralized BGP Egress Peer Engineering", draft-ietf-spring-segment-routing-central-epe-10 (work in progress), December 2017.

[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.ietf-spring-srv6-network-programming] Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-ietf-spring-srv6-network-programming-16 (work in progress), June 2020.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <https://www.rfc-editor.org/info/rfc2474>.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <https://www.rfc-editor.org/info/rfc2475>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[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>.

[RFC5439] Yasukawa, S., Farrel, A., and O. Komolafe, "An Analysis of Scaling Issues in MPLS-TE Core Networks", RFC 5439, DOI 10.17487/RFC5439, February 2009, <https://www.rfc-editor.org/info/rfc5439>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[RFC7471] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", RFC 7471, DOI 10.17487/RFC7471, March 2015, <https://www.rfc-editor.org/info/rfc7471>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC7810] Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 7810, DOI 10.17487/RFC7810, May 2016, <https://www.rfc-editor.org/info/rfc7810>.

[RFC7950] Bjorklund, M., Ed., "The YANG 1.1 Data Modeling Language", RFC 7950, DOI 10.17487/RFC7950, August 2016, <https://www.rfc-editor.org/info/rfc7950>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[RFC8571] Ginsberg, L., Ed., Previdi, S., Wu, Q., Tantsura, J., and C. Filsfils, "BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions", RFC 8571, DOI 10.17487/RFC8571, March 2019, <https://www.rfc-editor.org/info/rfc8571>.

Authors' Addresses

Jie Dong
Huawei Technologies

Email: jie.dong@huawei.com

Stewart Bryant
Futurewei Technologies

Email: stewart.bryant@gmail.com

Takuya Miyasaka
KDDI Corporation

Email: ta-miyasaka@kddi.com

Yongqing Zhu
China Telecom

Email: zhuyq8@chinatelecom.cn

Fengwei Qin
China Mobile

Email: qinfengwei@chinamobile.com

Zhenqiang Li
China Mobile

Email: li_zhenqiang@hotmail.com

Francois Clad
Cisco Systems

Email: fclad@cisco.com