SPRING
Internet-Draft
Intended status: Standards Track
Expires: January 13, 2021

S. Hegde
C. Bowers
Juniper Networks Inc.
X. Xu
Alibaba Inc.
A. Gulko
Refinitiv

July 12, 2020

Seamless Segment Routing
draft-hegde-spring-mpls-seamless-sr-00

Abstract

In order to operate networks with large numbers of devices, network operators organize networks into multiple smaller network domains. Each network domain typically runs an IGP which has complete visibility within its own domain, but limited visibility outside of its domain. Seamless Segment Routing (Seamless SR) provides flexible, scalable and reliable end-to-end connectivity for services across independent network domains. Seamless SR accomodates domains using SR, LDP, and RSVP for MPLS label distribution as well as domains running IP without MPLS (IP-Fabric).

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/.

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 January 13, 2021.

Table of Contents

1. Introduction

The Seamless SR architecture builds upon the Seamless MPLS architecture, which has been widely deployed to provide end-to-end transport for service in 3G/4G networks. [I-D.ietf-mpls-seamless-mpls], contains a good description of the Seamless MPLS architecture. Although [I-D.ietf-mpls-seamless-mpls] has not been published as an RFC, it serves as a useful description of the Seamless MPLS architecture. [I-D.ietf-mpls-seamless-mpls] describes the Seamless MPLS architecture, which uses LDP and/or RSVP for intra-domain label distribution, and BGP-LU [RFC3107] for end-to-end label distribution. The Seamless SR architecture builds on the the Seamless MPLS architecture. Seamless SR focuses on using segment routing for intra-domain label distribution.

By using segment routing for intra-domain label distribution, Seamless SR is able to easily support both SR-MPLS on IPv4 and IPv6 networks. This overcomes a limitation of the classic Seamless MPLS architecture, which was limited to run MPLS on IPv4 networks in practice. Seamless SR (like Seamless MPLS) can use BGP-LU (RFC 3107) to stitch different domains. However, Seamless SR can also take advantage of BGP Prefix-SID [RFC8669] to provide predictable and deterministic labels for the inter-domain connectivity.

5G technology is expected to place new requirements on the packet transport networks that support it. To enable 5G technology, packet transport networks will need to be capable of handling much greater bandwidth than today's 3G/4G networks. 5G networks are expected to require up to 250Gbps in the fronthaul and up to 400Gbps in the backhaul. The number of transport network devices is also expected to grow significantly to cater to 5G needs. Overall service availabilty requirements for 5G will place significant requirements on the resiliency of packet transport networks.

There is a desire to allow many 5G network functions to be virtualized and cloud native. In order to support latency-sensitive cloud-native 5G network functions, packet transport networks should be capable of providing low-latency paths end-to-end. Some services will require low-latency paths while others may require different QoS properties. The network should be able to differentiate the services and provide corresponding SLA transport paths.

The basic functionality of the Seamless SR architecture does not require any enhancements to existing protocols. However, in order to support end-to-end service requirements across multiple domains, protocol extensions may be needed. This draft discusses usecases, requirements, and potential protocol enhancements.

2. Terminology

This document uses the following terminology

- Access Node (AN): An access node is a node which processes customers frames or packets at Layer 2 or above. This includes but is not limited to DSLAMs and Cell Site Routers in 5G networks. Access nodes have only limited MPLS functionalities in order to reduce complexity in the access network.

- Pre-Aggregation Node (P-AGG): A pre-aggregation node (P-AGG) is a node which aggregates several access nodes (ANs).

- Aggregation Node (AGG): A aggregation node (AGG) is a node which aggregates several pre-aggregation nodes (P-AGG).

- Area Border Router (ABR): Router between aggregation and core domain.

- Label Switch Router (LSR): Label Switch router are pure transit nodes. ideally have no customer or service state and are therefore decoupled from service creation.

- Use Case: Describes a typical network including service creation points and distribution of remote node loopback prefixes.

Figure 1: Terminology

3. Use Cases

3.1. Service provider network

Service provider transport networks use multiple domains to support scalability. For this analysis, we consider a representative network design with four level of hierarchy: access domains, pre-aggregation domains, aggregation domains and a core. (See Figure 2). The 5G transport networks in particular are expected to scale to very large number of access nodes due to the shorter range of the 5G radio technology. The networks are expected to scale up to one million nodes.

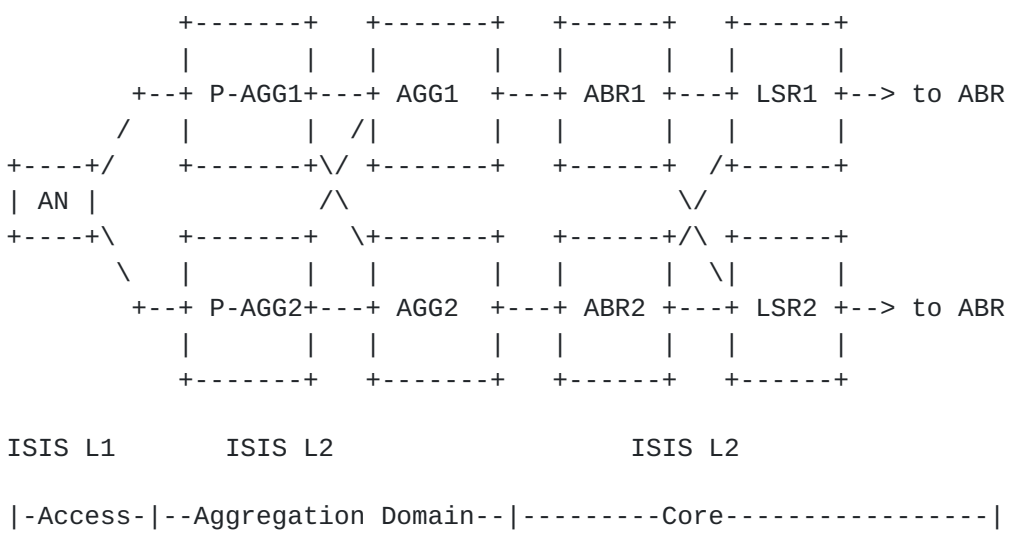

Figure 2: 5G network

Many network functions in a 5G network will be virtualized and distributed across multiple data centers. Virtualized network functions are instantiated dynamically across different compute resources. This requires that the underlying transport network supports the stringent SLA on end-to-end paths.

5G networks support variety of service use cases that require end-to-end slicing. In certain cases the end-to-end connectivity requires differentiated forwarding capabilities. Seamless SR architecture should provide ability to establish end-to-end paths that satisfy the required SLAs. For Example, End user requirement could be to establish low latency path end-to-end. The System Architecture for the 5G System [TS.23.501-3GPP] currently defines four standardized Slice/Service Types: Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communication (URLLC), massive Internet of Things (mIoT), Vehicle to everything (V2X). The Seamless SR should support end-to-

end QoS mechansisms to allow the creation of network slices with these four Slice/Service Types.

Many deployments consist of ring topologies in the access and aggregation networks. In the ring topologies, there are atmost two forwarding paths for the traffic, where as the core networks consist of nodes with more denser connectivity compared to ring topologies. Thus core networks may have larger number of TE paths while access networks will have smaller number of TE paths. The Seamless SR architecture should support ability to have more TE paths in one domain and lesser number of TE paths in another domain and provide ability to effectively connect the domain end-to-end satisfying end-to-end constraints.

3.2. Large scale WAN networks

As WAN networks grow beyond several thousand nodes, it is often useful to divide the network into multiple IGP domains. The different IGP domains provide better fault isolation. Smaller IGP domains can also reduce FIB scale.

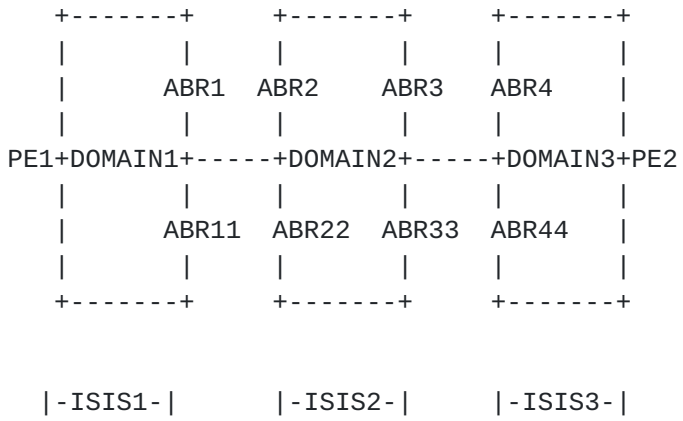

Figure 3: WAN Network

Large WAN networks often cross national boundaries. In order to meet data sovereignity requirements, operators need to maintain strict control over end-to-end traffic-engineered(TE) paths. Segment Routing provides two main solutions to implement highly constrained TE paths. Flex-algo (defined in [I-D.ietf-lsr-flex-algo]) uses prefix-SIDs computed by all nodes in the IGP domain using the same pruned topology. Highly constrained TE paths for the data soveriegnty use case can also be implemented using SR-TE policies ([I-D.ietf-spring-segment-routing-policy]) built using unprotected adjacency SIDs.

Both of these approaches work well for intra-domain TE paths. However, they both have limitations when one tries to extend them to the creation of highly constrained inter-domain TE paths. A goal of seamless SR is to be able to create highly constrained inter-domain TE paths in a scalable manner.

3.3. Data Center Interconnect (DCI) Networks

Data centers are playing an increasingly important role in providing access to information and applications. Geographically diverse data centers usually connect via a high speed, reliable and secure core network.

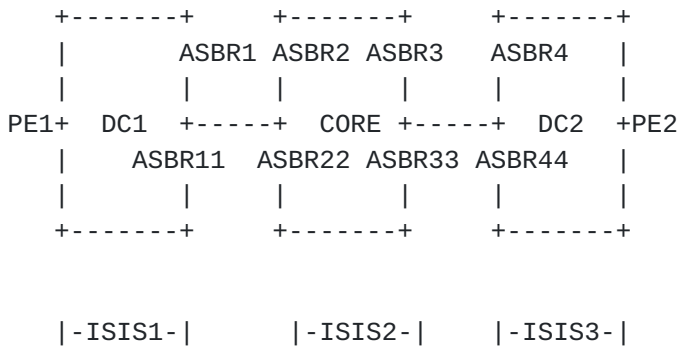

Figure 4: DCI Network

In many Data Center deployments, applications require end-to-end path diversity and/or end-to-end low latency paths. It is desirable to have a uniform technology deployed in the core as well as in the Data Centers to create these SLA paths. Such uniformity simplifies the network to a great extent. It is desirable for a solution to only require service-related configurations on the access end-points where services are attached, avoiding service-related configurations on the ABR/ASBR nodes.

3.4. Multicast Usecases

Multicast services such as IPTV and multicast also need to be support across a multi-domain service provider network. Multicast services such as IPTV, multicast VPN etc need to be supported in a service provider network.

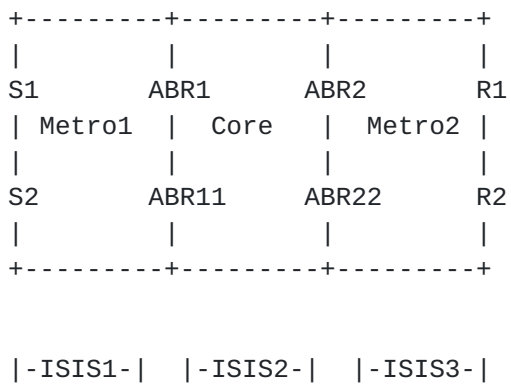

Figure 5: Multicast usecases

Figure 5 shows a simplified multi-domain network supporting multicast. Multicast sources S1 and S2 lie in a different domain from the receivers R1 and R2. Using multiple IGP domains presents a problem for the establishment of multicast replication trees. Typically, a multicast receiver does a reverse path forwarding (RPF) lookup for a multicast source. One solution is to leak the routes for multicast sources across the IGP domains. However, this can compromise the scaling properties of the multi-domain architecture. SR-P2MP [I-D.voyer-pim-sr-p2mp-policy] offers a solution for both intra-domain and inter-domain multicast. However, it does accomodate deployments using existing intra-domain multicast technology, such as mLDP [RFC6388] in some of the domains. A solution should accomodate a mixture of existing and newer technologies to better facilitate coexistence and migration.

4. Requirements

This section provides a summary of requirements derived from the use cases described in previous sections.

4.1. MPLS Transport

The architecture should provide MPLS transport between two service endpoints regardless of whether the two end-points are in the same IGP domain, different IGP domains, or in different autonomous systems.

The MPLS transport should be supported on IPv4, IPv6, and dual-stack networks.

4.2. SLA Guarantee

The architecture should allow the creation of paths that support end-to-end SLAs. The paths should for example obey constraints related to latency, diversity, and availability.

The architecture should support end-to-end network slicing as described by 5G transport requirements [TS.23.501-3GPP].

4.3. Scalability

The architecture should be able to support up to 1 million nodes.

The architecture should facilitate the use of access nodes with low RIB/FIB and low CPU capabilities.

The architecture should facilitate the use of access nodes with low label stacking capability.

The architecture should allow for a scalable response to network events. An individual node should only need to respond to a limited subset of network events.

Service routes on the border nodes should be minimized.

4.4. Availability

Traffic should be Fast Reroute (FRR) protected against link, node, and SRLG failures within a domain.

Traffic should be Fast Reroute (FRR) protected against border node failures.

Traffic should be Fast Reroute (FRR) protected against egress node and egress link failures.

4.5. Operations

Each domain should be independent and should not depend on the transport technology in another domain. This allows for more flexible evolution of the network.

Basic MPLS OAM mechanisms described in [RFC8029] should be supported.

End-to-end mpls ping and traceroute procedures should be supported.

Ability to validate the path inside each domain should be supported.

Statistics for inter-domain paths on the ingress and egress PE nodes as well as border nodes should be supported.

4.6. Service Mapping

The architecture should support the automated steering of traffic on to transport paths based on communities carried in the service prefix advertisements.

The architecture should support the steering of traffic on to transport paths based the DSCP value carried in IPv4/IPv6 packets.

Traffic steering based on EXP bits in the mpls header should be supported.

Traffic steering based on 5-tuple packet filter should be supported. Source address, destination address, source port, destination port and protocol fields should be allowed.

All traffic steering mechanims should be supported for all kinds of service traffic including VPN traffic as well as global internet traffic.

The core domain is expected to have more traffic enginnering constraints as compared to metros. The ability to map the services to appropriate transport tunnels at service attachment points should be supported.

5. Seamless Segment Routing architecture

5.1. Solution Concepts

The solution described below makes use of the following concepts.

- Transport Class (TC): A Transport Class is defined as a collection of end-to-end MPLS paths that satisfy a set of constraints or Service Level Aggreements.

- BGP-Classful Transport (BGP-CT): A new BGP family used to establish Transport Class paths across different domains.

- Route Distinguisher (RD): The Route Distinguisher is defined in RFC4364. In BGP-CT, the RD is used in BGP advertisements to differentiate multiple paths to the same loopback address. It may be useful to automatically generate RDs in order to simplify configuration.

- Route Target (RT): The Route Target extended community is carried in BGP-CT advertisements. The RT represents the Transport Class of an advertised path.

- Mapping Community (MC): The Mapping Community is the standard BGP community as defined in RFC1997. In the Seamless SR architecture, an MC is carried by a service route. The MC is used to identify the specific local policy used to map traffic for a service route to different Transport Class paths. The local policy can include additional traffic steering properties for placing traffic on different Transport Class paths. The values of the MCs and the corresponding local policies for service mapping are defined by the network operator.

Figure 6: Solution Concepts

5.2. BGP Classful Transport

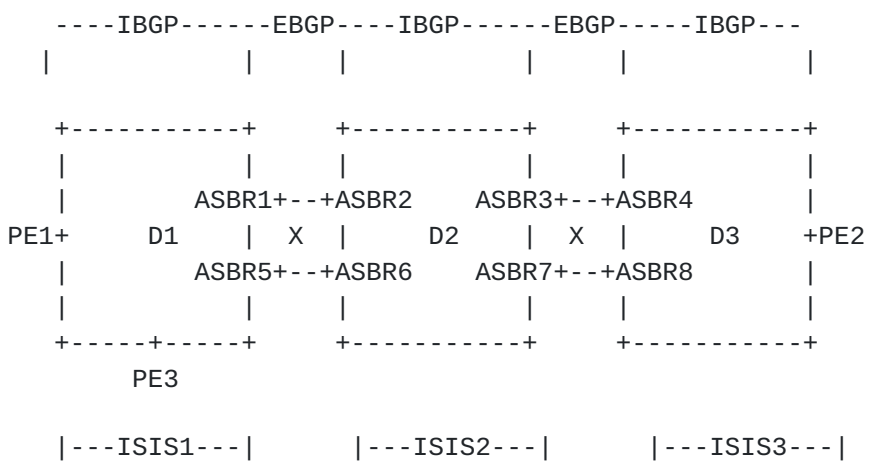

Figure 7: WAN Network

The above diagram shows a WAN network divided into 3 different domains. Within each domain, BGP sessions are established between the PE nodes and the border nodes as well as between border nodes. BGP sessions are also established between border nodes across domains. The goal is for PE1 to have MPLS connectivity to PE2, satisfying specific characteristics. Multiple MPLS paths from PE1 to PE2 are required in order to satisfy diffrent SLAs. [I-D.kaliraj-idr-bgp-classful-transport-planes] defines a new BGP family called BGP-Classful Transport. The NLRI for this new family consists of a prefix and a Route Distinguisher. The prefix corresponds to the loopback of the destination PE, and RD is used to distinguish multiple paths to the same PE loopback. The BGP-CT advertisement also carries a Route Target. The RT specifies the Transport Class to which the BGP-CT advertisement belongs.

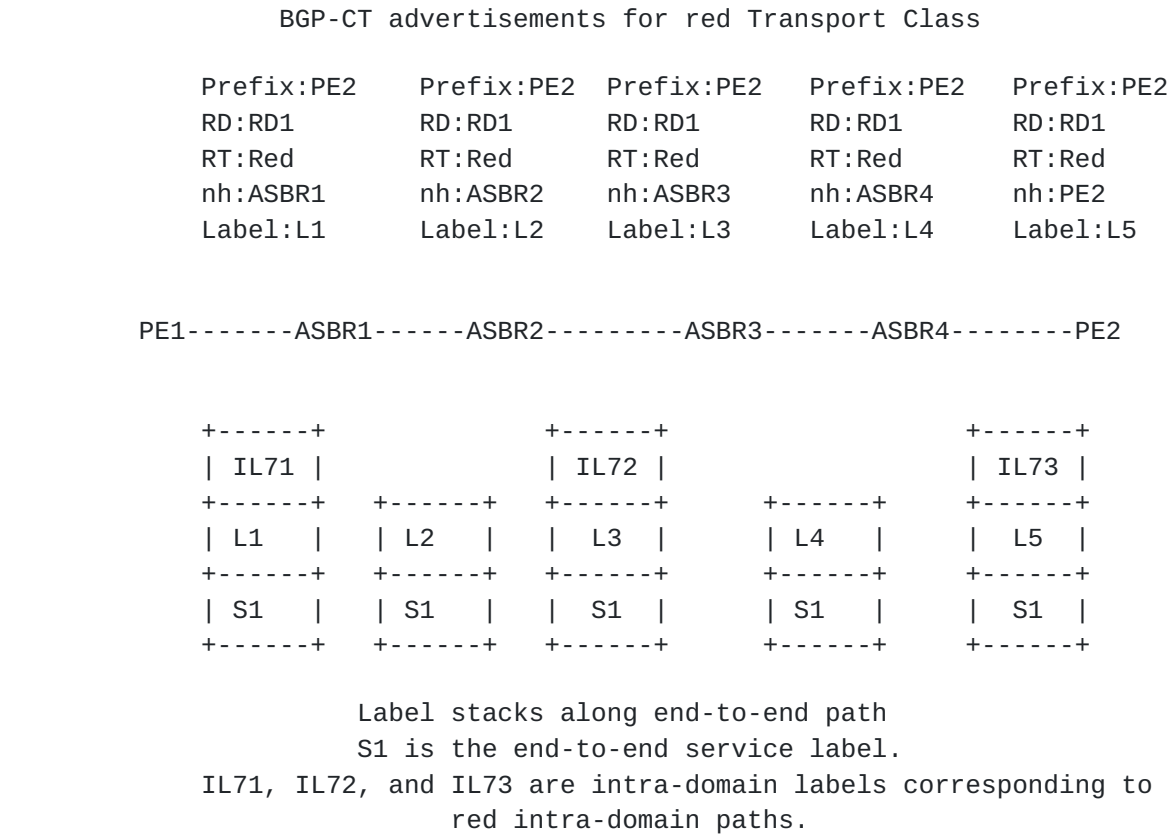

Figure 8: BGP-CT Advertisements and Label Stacks

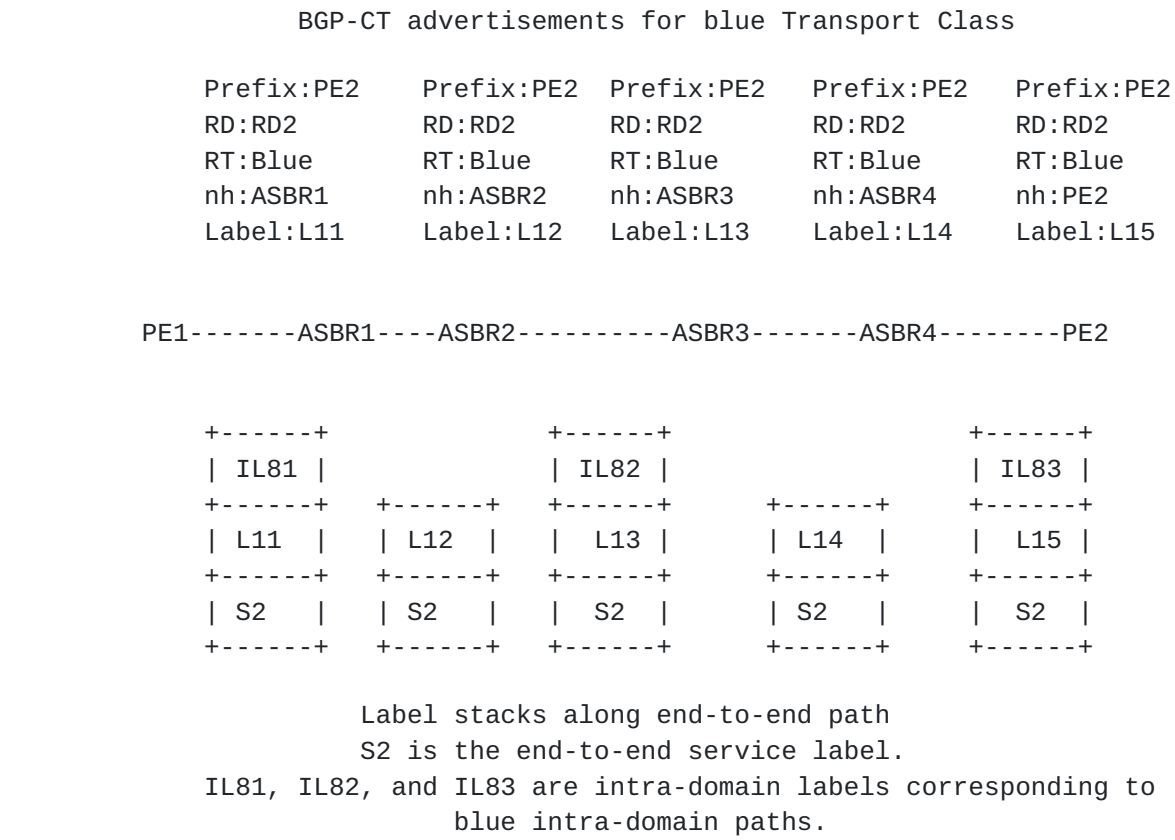

Figure 9: BGP-CT Advertisements and Label Stacks

For example, consider the diagram in Figure 8 and Figure 9 . The diagram shows the BGP-CT advertisements corresponding to two different end-to-end paths between PE1 and PE2. The two different paths belong to two different Transport Classes, red and blue. In order to create unique NLRIs for the two advertisements, PE2 uses two different RDs. In the example above, the red BGP-CT advertisement has an RD of RD1 and the blue BGP-CT advertisement has an RD of RD2. The advertisements will have RTs corresponding to the red and blue Transport Classes respectively. The RT MAY be directly mapped from the color extended community defined in [I-D.ietf-idr-tunnel-encaps]. In addition to the red and blue BGP-CT advertisments, the diagram shows the label stacks at different points along the end-to-end paths for the forwarding entries which are established by the two advertisements. Labels L1-L4 are red BGP-CT labels advertised by border nodes ASBR1,2,3,and 4, while label L5 is advertised by PE2 for the red Transport Class. Labels L11-L14 are blue BGP-CT labels advertised by border nodes ASBR1,2,3,and 4, while label L15 is advertised by PE2 for the blue Transport Class.

IL71, IL72, and IL73 represent tunnels internal to the domains 1, 2, and 3 which correspond to the red Transport Class. IL81, IL82, and IL83 represent tunnels internal to the domains 1, 2, and 3 which correspond to the blue Transport Class. In this example, we assume that the intra-domain tunnels correspond to SRTE policies having red SRTE-policy-color and blue SRTE-policy-color. Service labels are represented by S1 and S2. In this example, we assume that the service advertisement corresponding to S1 carries the red extended-color community, while the service advertisement corresponding to S2 carries the blue extended-color community. By default, the Transport Class carried in the BGP-CT route target maps to the extend-color community as well as the SRTE-policy-color. Therefore, based on the simple BGP-CT advertisment originated by PE2, PE1 is able to automatically steer traffic for service S1 over an end-to-end path made up of red SRTE policies in each domain.

Note that this example focuses on how signalling originated by PE2 results in forwarding state used by PE1 to reach PE2 on a specific Transport Class path. The solution supports the establishment of forwarding state for an arbitrary number of PEs to reach PE2. For example, PE3 in Figure 8 can reach PE2 on a red Transport Class path established using the same BGP-CT signalling. The signalling and forwarding state from ASBR1 all the way to PE2 is common to the paths used by both PE1 and PE3. This merging of signalling and forwarding state is essentially to the good scaling properties of the Seamless SR architecture. Millions of end-to-end Transport Class paths can be established in a scalable manner.

5.3. SLA Guarantee

5.3.1. Low latency

In a 5G network, many network functions are virtualized and distributed. Certain functions are time and latency sensitive. Latency is one of the main SLA parameter for 5G networks. In inter-domain networks, End-to-End latency measurement is required. Inside a domain, latency measurement mechanisms such as TWAMP [RFC5357] are used and link latency is advertised in IGP using extensions described in [RFC8570]and [RFC7471] .

[I-D.ietf-idr-performance-routing] extends the BGP AIGP attribute [RFC7311] by adding a sub TLV to carry an accumulated latency metric. The BGP best path selection algorithm used for a Transport Class requiring low latency will consider the accumulated latency metric to choose lowest latency path.

5.3.2. Traffic Engineering (TE) constraints

TE constraints generally include the ability to send traffic via certain nodes or links or avoid using certain nodes or links. In the Seamless SR architecture, the intra-domain transport technology is responsible for ensuring the TE constraints inside the domain, BGP-CT ensures that the end-to-end path is construct from intra-domain paths and inter-AS links that individually satisfy the TE constraints.

For example, in order to construct a pair of diverse paths, we can define a red and a blue Transport Class. Within each domain, the red and blue Transport Class path are realized using intra-domain path diversity mechansisms. For example, in a domain using flex-algo, red and blue Transport Classes are realized using red and blue flex-algo which don't share any links. To maintain path diversity on inter-AS links, BGP policies are used to associate two inter-AS peers with the red Transport Class and another two inter-AS peers with the blue Transport Class.

5.3.3. Bandwidth constraints

The Seamless SR architecture does not natively support end-to-end bandwidth reservations. In this architecture, the bandwidth utilization characteristics of each domain are managed independently. The intra-domain bandwidth management can make use of a variety of tools.

Link bandwidth extended community as defined in [I-D.ietf-idr-link-bandwidth] allows for efficient weighted load-balancing of traffic on multiple BGP-CT paths that belong to the same Transport Class. For optimized path placement, a seperate tool may be deployed and BGP policies/communites used for path placement.

5.4. Scalability

5.4.1. Access node scalability

The Seamless SR architecture needs to be able to accommodate very large numbers of access devices. These access devices are expected to be low-end devices with limited FIB capacity. The Seamless MPLS architecture, as described in [I-D.ietf-mpls-seamless-mpls], recommends the use of LDP DOD mode to limit the size of both the RIB and the FIB needed on the access devices. In the Seamless SR architecture, networks use IGP based label distribution and do not have this selective label request mechanism. However, RIB scalability of access nodes has not been a problem for real seamless MPLS deployments. In cases where access devices are low on CPU and memory and unable to support large a RIB, BGP filtering policies can

be applied at the ABR/ASBR routers to restrict the number of BGP-CT advertisements towards the access devices. The access devices will receive only the PE loopbacks that it needs to connect to.

5.4.2. Label stack depth

The ability for a device to push multiple MPLS labels on a packet depends on hardware capabilities. Access devices are expected to have limited label stack push capabilities. The Seamless SR architecture can provide cross-domain MPLS connectivity with a single label. The access devices push one service label, one BGP-CT label, and one intra-domain transport label. Assuming shortest path SR-MPLS in the access domain, the access domain transport will use a single label. Light weight traffic-engineering and slicing could also be achieved with a single label as described in [I-D.ietf-lsr-flex-algo]. The access nodes will need to be able to push a minimum of 3 labels.

5.4.3. Label Resources

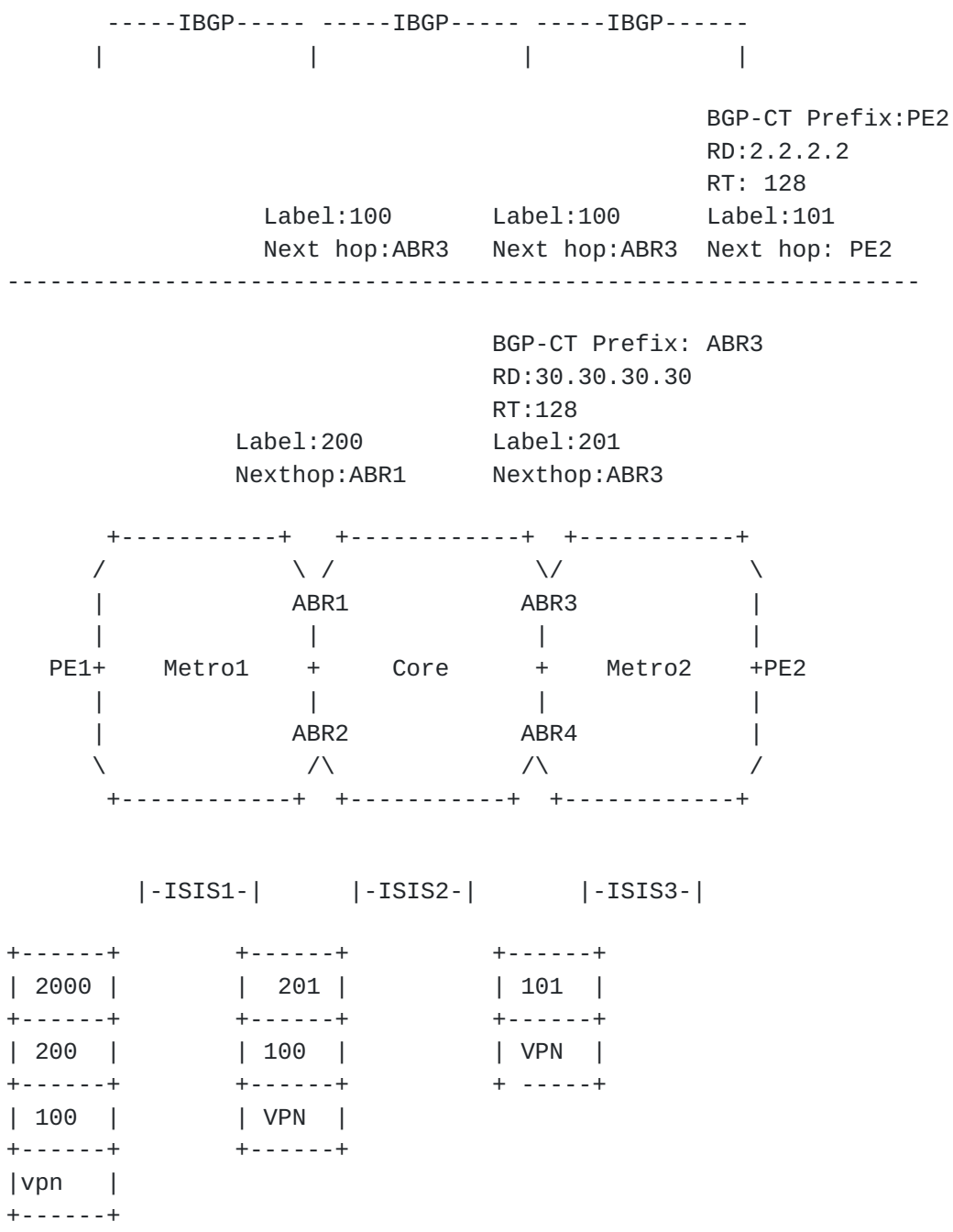

Figure 10: Recursive Route Resolution

The label resources are an important consideration in MPLS networks. On access devices, labels are consumed by services as well as for transport loopbacks inside IGP domain where the access device resides. For example, in the above diagram PE1 would have to allocate label resources equal to the number of customers connecting (i.e. the number of L2/L3 VPNs). Based on the size of the IGP domain

that PE1 resides in, it will also have to allocate labels for IGP loopbacks. This number is at most a few thousands. So overall a typical access device should have adequate label resources in Seamless SR architecture. The P routers need to allocate labels for IGP loopbacks. This number again is small. At most it will be a few thousand based on number of nodes in the largest IGP domains. The metro networks connect to the core network through ABRs. It is possible that a given ABR may end up having to maintain forwarding entries for a large subset of the transport loopback routes. There may be a large number of metro networks connecting to a given ABR, and in this case, the ABR will need forwarding entries for every access node in the directly connected metros. So, this ABR may have to maintain on the order of 100k routes. With BGP-CT each Transport Class will have to be separately allocated a label. So, in the above example, the ABR1 would have to use 300k labels if there were 3 Transport Classes. MPLS labels are 20 bit long and the label range of 16-1 million is available for general applications.This label space is shared between transport protocols and services. However, in a well-designed network, ABRs are not expected to host service routes. This leaves with 1 million labels completely available for transport infrastructure. This is sufficient in most cases.

In certain cases, it is desirable to reduce the forwarding state on the ABRs. This reduction can be achieved with label stacking as a result of recursive route resolution. In the Figure 10, PE2 advertises a BGP-CT prefix with nexthop being PE2 and 101 label. ABR3 advertises a label 100 for this BGP-CT prefix and changes the nexthop to self. When ABR1 receives this BGP-CT advertisement for PE2, it does not change the nexthop and advertises same label advertised by ABR3. When PE1 receives the BGP-CT advetisement for PE2 with a nexthop of ABR3, it resolves on another BGP-CT prefix for ABR3. As shown in the diagram, ABR3 advertises BGP-CT prefix with 201 and ABR1 advertises label 200 and sets nexthop to self. On PE1, the data packet consists of a VPN label at the bottom followed by 2 BGP-CT labels 100 and 200. The top most label 2000 is the transport label for the metro1 domain. There is 1 additional BGP-CT label on the datapacket.

Recursive route resolution provides significant forwarding state reduction on the ABRs. ABRs have to allocate label resources for the PE loopback that they directly connect to. This number is significantly lower as compared to the total number of PEs in the network.

5.5. Reliability

Transport layer redundancy is very important in 5G networks. Any link or node failure must be repaired with 50ms conevergence time. 50 ms convergence time can be achieved with Fast ReRoute (FRR) mechanisms. Seamless SR architecture supports Intra-domain link/node failures, Border node failures and the egress node and link failures for 50 ms convergence. Details of the FRR techniques are described in below sections.

5.5.1. Intra domain link and node protection

In the seamless SR architecture, protection against node and link failure is achieved with the relevant FRR techniques for the corresponding transport mechanism used inside the domain. In the case of an IP fabric, ECMP FRR or LFA can be used. In SR networks, TI-LFA [I-D.ietf-rtgwg-segment-routing-ti-lfa] provides link and node protection. For SR-TE [I-D.ietf-spring-segment-routing-policy] transport, link and node protection can be achieved using TI-LFA, combined with mechanisms described in [I-D.hegde-spring-node-protection-for-sr-te-paths].

5.5.2. Egress Link and node protection

[RFC8679] describes the mechanisms for providing protection for border nodes and PE devices where services are hosted. The mechanism can be further simplified operationally with anycast SIDs and anycast service labels, as described in [I-D.hegde-rtgwg-egress-protection-sr-networks].

5.5.3. Border Node protection

Border node protection is very important in a network consisting of multiple domains. Seamless SR architecture proposes to achieve 50ms FRR protection in the event of node failure with anycast address for the ABR/ASBRs and allocates same label for the BGP-CT Prefix.The detailed mechanism is described in [I-D.hegde-rtgwg-egress-protection-sr-networks].

5.6. Operations

5.6.1. MPLS ping and Traceroute

Seamless SR Architecture is based on hierarchical network modeling. The End-to-end BGP-CT connectivity can be verified. A new FEC is defined for BGP-CT as defined in draft [I-D.kaliraj-idr-bgp-classful-transport-planes] that describes End-to-End connectivity verification as well as fault isolation. The

BGP-CT verification happens only on the BGP nodes. The intra-domain connectivity verification and fault isolation will be based on the technology deployed in that domain as defined in [RFC8029] and [RFC8287].

5.6.2. Counters and Statistics

Traffic accounting and ability to build demand matrix for PE to PE traffic is very important. With BGP-CT, per-label transit counters should be supported on every transit router. per-label transit counters provide details of total traffic towards a remote PE measured at every BGP transit router. per-label egress counter should be supported on ingress PE router. per-label egress counter provides total traffic from ingress PE to the specific remote PE.

5.7. Service Mapping

Service mapping is an imprtant aspect of any architecture. It provides means to translate end users SLA requirements into operator's network configurations. Seamless SR architecture supports automatic steering with extended color community. The Transport Class and the route target carried by the BGP-CT advertisement directly map to the extended color community. Services that require specific SLA carry the extended color community which maps to the Transport Class to which the BGP-CT advertisement belongs.

Other types of traffic steering such as DSCP based forwarding is expressed with mapping-community. Mapping community is a standard BGP community and is completely generic and user defined. The mapping community will have a specific service mapping feature associated with it along with required fallback behaviour when the primary transport goes down. The below list provides a general guideline into the different service mapping features and fallback options an implementation should provide.

DSCP based mapping with each DSCP mapping to a Transport Class.

DSCP based mapping with default mapping to a best-effort transport

DSCP based mapping with fallback to best-effort when primary transport tunnel goes down.

Extended color community based mapping with fallback to best effort

Fallback options with specific protocol during migrations

Falback options to a different Transport Class.

No Fallback permitted.

5.8. Migrations

Networks that migrate from Seamless MPLS architecture to Seamless SR architecture, require that all the border nodes and PE devices be upgraded and enable new family on the BGP session. In cases where legacy nodes that cannot be upgraded exporting from BGP-LU into BGP-CT and vice versa SHOULD be supported.

5.9. Interworking with v6 transport technologies

A later version of this document will address interworking with other v6 technologies, including SRv6, SRm6, and MPLS over GRE6.

5.10. BGP based Multicast

BGP based multicast as described in draft [I-D.zzhang-bess-bgp-multicast] serves two main purposes. It can replace PIM/ mLDP inside a domain to natively do a BGP based multicast. It can also serve as an overlay stitching protocol to stitch multiple P2MP LSPs across the domain. This gives the ability to easily transition each domain independently from one technology to the other. BGP based multicast defines a new SAFI for carrying the MULTICAST TREE SAFI. Different route types are defined to support the various usecases.

6. Backward Compatibility

7. Security Considerations

TBD

8. IANA Considerations

9. Acknowledgements

Many thanks to Kireeti Kompella, Ron Bonica, Krzysztof Szarcowitz, Srihari Salngi,Julian Lucek for discussions and inputs.

10. Contributors

1.Kaliraj Vairavakkalai

Juniper Networks

kaliraj@juniper.net

2. Jeffrey Zhang

Juniper Networks

zzhang@juniper.net

11. References

11.1. Normative References

[I-D.hegde-rtgwg-egress-protection-sr-networks]
Hegde, S. and W. Lin, "Egress Protection for Segment Routing (SR) networks", draft-hegde-rtgwg-egress-protection-sr-networks-00 (work in progress), March 2020.

[I-D.ietf-idr-performance-routing]
Xu, X., Hegde, S., Talaulikar, K., Boucadair, M., and C. Jacquenet, "Performance-based BGP Routing Mechanism", draft-ietf-idr-performance-routing-02 (work in progress), October 2019.

[I-D.kaliraj-idr-bgp-classful-transport-planes]
Vairavakkalai, K., Venkataraman, N., and B. Rajagopalan, "BGP Classful Transport Planes", draft-kaliraj-idr-bgp-classful-transport-planes-00 (work in progress), May 2020.

[I-D.zzhang-bess-bgp-multicast]
Zhang, Z., Giuliano, L., Patel, K., Wijnands, I., mishra, m., and A. Gulko, "BGP Based Multicast", draft-zzhang-bess-bgp-multicast-03 (work in progress), October 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>.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, DOI 10.17487/RFC3107, May 2001, <https://www.rfc-editor.org/info/rfc3107>.

[RFC8669] Previdi, S., Filsfils, C., Lindem, A., Ed., Sreekantiah, A., and H. Gredler, "Segment Routing Prefix Segment Identifier Extensions for BGP", RFC 8669, DOI 10.17487/RFC8669, December 2019, <https://www.rfc-editor.org/info/rfc8669>.

11.2. Informative References

[I-D.hegde-spring-node-protection-for-sr-te-paths]
Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu, "Node Protection for SR-TE Paths", draft-hegde-spring-node-protection-for-sr-te-paths-05 (work in progress), July 2019.

[I-D.ietf-idr-link-bandwidth]
Mohapatra, P. and R. Fernando, "BGP Link Bandwidth Extended Community", draft-ietf-idr-link-bandwidth-07 (work in progress), March 2018.

[I-D.ietf-idr-tunnel-encaps]
Patel, K., Velde, G., and S. Ramachandra, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-15 (work in progress), December 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-mpls-seamless-mpls]
Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz, M., and D. Steinberg, "Seamless MPLS Architecture", draft-ietf-mpls-seamless-mpls-07 (work in progress), June 2014.

[I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B., Francois, P., Voyer, D., Clad, F., and P. Camarillo, "Topology Independent Fast Reroute using Segment Routing", draft-ietf-rtgwg-segment-routing-ti-lfa-03 (work in progress), March 2020.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-07 (work in progress), May 2020.

[I-D.voyer-pim-sr-p2mp-policy]
Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. Zhang, "Segment Routing Point-to-Multipoint Policy", draft-voyer-pim-sr-p2mp-policy-02 (work in progress), July 2020.

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, DOI 10.17487/RFC1997, August 1996, <https://www.rfc-editor.org/info/rfc1997>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <https://www.rfc-editor.org/info/rfc5357>.

[RFC6388] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, DOI 10.17487/RFC6388, November 2011, <https://www.rfc-editor.org/info/rfc6388>.

[RFC7311] Mohapatra, P., Fernando, R., Rosen, E., and J. Uttaro, "The Accumulated IGP Metric Attribute for BGP", RFC 7311, DOI 10.17487/RFC7311, August 2014, <https://www.rfc-editor.org/info/rfc7311>.

[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>.

[RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N., Aldrin, S., and M. Chen, "Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures", RFC 8029, DOI 10.17487/RFC8029, March 2017, <https://www.rfc-editor.org/info/rfc8029>.

[RFC8287] Kumar, N., Ed., Pignataro, C., Ed., Swallow, G., Akiya, N., Kini, S., and M. Chen, "Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with MPLS Data Planes", RFC 8287, DOI 10.17487/RFC8287, December 2017, <https://www.rfc-editor.org/info/rfc8287>.

[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>.

[RFC8570] Ginsberg, L., Ed., Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 8570, DOI 10.17487/RFC8570, March 2019, <https://www.rfc-editor.org/info/rfc8570>.

[RFC8679] Shen, Y., Jeganathan, M., Decraene, B., Gredler, H., Michel, C., and H. Chen, "MPLS Egress Protection Framework", RFC 8679, DOI 10.17487/RFC8679, December 2019, <https://www.rfc-editor.org/info/rfc8679>.

[TS.23.501-3GPP]
3rd Generation Partnership Project (3GPP), "System Architecture for 5G System; Stage 2, 3GPP TS 23.501 v16.4.0", March 2020.

Authors' Addresses

Shraddha Hegde
Juniper Networks Inc.
Exora Business Park
Bangalore, KA 560103
India

Email: shraddha@juniper.net

Chris Bowers
Juniper Networks Inc.

Email: cbowers@juniper.net

Xiaohu Xu
Alibaba Inc.
Beijing
China

Email: xiaohu.xxh@alibaba-inc.com

Arkadiy Gulko
Refinitiv

Email: arkadiy.gulko@refinitiv.com