TEAS Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2022

T. Saad
V. Beeram
Juniper Networks
B. Wen
Comcast
D. Ceccarelli
J. Halpern
Ericsson
S. Peng
R. Chen
ZTE Corporation
X. Liu
Volta Networks
L. Contreras
Telefonica
R. Rokui
Nokia

July 11, 2021

Realizing Network Slices in IP/MPLS Networks
draft-bestbar-teas-ns-packet-03

Abstract

Network slicing provides the ability to partition a physical network into multiple logical networks of varying sizes, structures, and functions so that each slice can be dedicated to specific services or customers. Network slices need to operate in parallel while providing slice elasticity in terms of network resource allocation. The Differentiated Service (Diffserv) model allows for carrying multiple services on top of a single physical network by relying on compliant nodes to apply specific forwarding treatment (scheduling and drop policy) on to packets that carry the respective Diffserv code point. This document proposes a solution based on the Diffserv model to realize network slicing in IP/MPLS networks.

Table of Contents

1. Introduction

Network slicing allows a Service Provider to create independent and logical networks on top of a common or shared physical network infrastructure. Such network slices can be offered to customers or used internally by the Service Provider to facilitate or enhance their service offerings. A Service Provider can also use network slicing to structure and organize the elements of its infrastructure. This document provides a path control technology agnostic solution that a Service Provider can deploy to realize network slicing in IP/MPLS networks.

The definition of network slice for use within the IETF and the characteristics of IETF network slice are specified in [I-D.ietf-teas-ietf-network-slice-definition]. A framework for reusing IETF VPN and traffic-engineering technologies to realize IETF network slices is discussed in [I-D.nsdt-teas-ns-framework]. These documents also discuss the function of an IETF Network Slice Controller and the requirements on its northbound and southbound interfaces.

This document introduces the notion of a slice aggregate which comprises of one of more IETF network slice traffic streams. It describes how a slice policy can be used to realize a slice aggregate by instantiating specific control and data plane behaviors on select topological elements in IP/MPLS networks. The onus is on the IETF Network Slice Controller to maintain the mapping between one or more IETF network slices and a slice aggregate. The mechanisms used by the controller to determine the mapping are outside the scope of this document. The focus of this document is on the mechanisms required at the device level to address the requirements of network slicing in packet networks.

In a Differentiated Service (Diffserv) domain [RFC2475], packets requiring the same forwarding treatment (scheduling and drop policy) are classified and marked with a Class Selector (CS) at domain ingress nodes. At transit nodes, the CS field inside the packet is inspected to determine the specific forwarding treatment to be

applied before the packet is forwarded further. Similar principles are adopted by this document to realize network slicing.

When logical networks representing slice aggregates are realized on top of a shared physical network infrastructure, it is important to steer traffic on the specific network resources allocated for the slice aggregate. In packet networks, the packets that traverse a specific slice aggregate MAY be identified by one or more specific fields carried within the packet. A slice policy ingress boundary node populates the respective field(s) in packets that enter a slice aggregate to allow interior slice policy nodes to identity those packets and apply the specific Per Hop Behavior (PHB) that is associated with the slice aggregate. The PHB defines the scheduling treatment and, in some cases, the packet drop probability.

The slice aggregate traffic may further carry a Diffserv CS to allow differentiation of forwarding treatments for packets within a slice aggregate. For example, when using MPLS as a dataplane, it is possible to identify packets belonging to the same slice aggregate by carrying a global MPLS label in the label stack that identifies the slice aggregate in each packet. Additional Diffserv classification may be indicated in the Traffic Class (TC) bits of the global MPLS label to allow further differentiation of forwarding treatments for traffic traversing the same slice aggregate network resources.

This document covers different modes of slice policy and discusses how each slice policy mode can ensure proper placement of slice aggregate paths and respective treatment of slice aggregate traffic.

1.1. Terminology

The reader is expected to be familiar with the terminology specified in [I-D.ietf-teas-ietf-network-slice-definition] and [I-D.nsdt-teas-ns-framework].

The following terminology is used in the document:

IETF network slice:
: a well-defined composite of a set of endpoints, the connectivity requirements between subsets of these endpoints, and associated requirements; the term 'network slice' in this document refers to 'IETF network slice' as defined in [I-D.ietf-teas-ietf-network-slice-definition].

IETF Network Slice Controller (NSC):
: controller that is used to realize an IETF network slice [I-D.ietf-teas-ietf-network-slice-definition].

Slice policy:
a policy construct that enables instantiation of mechanisms in support of IETF network slice specific control and data plane behaviors on select topological elements; the enforcement of a slice policy results in the creation of a slice aggregate.

Slice aggregate:
a collection of packets that match a slice policy selection criteria and are given the same forwarding treatment; a slice aggregate comprises of one or more IETF network slice traffic streams; the mapping of one or more IETF network slices to a slice aggregate is maintained by the IETF Network Slice Controller.

Slice policy capable node:
a node that supports one of the slice policy modes described in this document.

Slice policy incapable node:
a node that does not support any of the slice policy modes described in this document.

Slice aggregate traffic:
traffic that is forwarded over network resources associated with a specific slice aggregate.

Slice aggregate path:
a path that is setup over network resources associated with a specific slice aggregate.

Slice aggregate packet:
a packet that traverses network resources associated with a specific slice aggregate.

Slice policy topology:
a set of topological elements associated with a slice policy.

Slice aggregate aware TE:
a mechanism for TE path selection that takes into account the available network resources associated with a specific slice aggregate.

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.

1.2. Acronyms and Abbreviations

BA: Behavior Aggregate

CS: Class Selector

SS: Slice Selector

S-PHB: Slice policy Per Hop Behavior as described in Section 5.1.3

SSL: Slice Selector Label as described in Section 5.1.1

SSLI: Slice Selector Label Indicator

SLA: Service Level Agreement

SLO: Service Level Objective

Diffserv: Differentiated Services

MPLS: Multiprotocol Label Switching

LSP: Label Switched Path

RSVP: Resource Reservation Protocol

TE: Traffic Engineering

SR: Segment Routing

VRF: VPN Routing and Forwarding

2. Network Resource Slicing Membership

A slice aggregate can be instantiated over parts of an IP/MPLS network (e.g., all or specific network resources in the access, aggregation, or core network), and can stretch across multiple domains administered by a provider. A slice policy topology may include all or a sub-set of the physical nodes and links of an IP/MPLS network; it may be comprised of dedicated and/or shared network resources (e.g., in terms of processing power, storage, and bandwidth).

2.1. Dedicated Network Resources

Physical network resources may be fully dedicated to a specific slice aggregate. For example, traffic belonging to a slice aggregate can traverse dedicated network resources without being subjected to

contention from traffic of other slice aggregates. Dedicated network resource slicing allows for simple partitioning of the physical network resources amongst slice aggregates without the need to distinguish packets traversing the dedicated network resources since only one slice aggregate traffic stream can traverse the dedicated resource at any time.

2.2. Shared Network Resources

To optimize network utilization, sharing of the physical network resources may be desirable. In such case, the same physical network resource capacity is divided among multiple slice aggregates. Shared network resources can be partitioned in the data plane (for example by applying hardware policers and shapers) and/or partitioned in the control plane by providing a logical representation of the physical link that has a subset of the network resources available to it.

3. Path Selection

Path selection in a network can be network state dependent, or network state independent as described in Section 5.1 of [I-D.ietf-teas-rfc3272bis]. The latter is the choice commonly used by IGPs when selecting a best path to a destination prefix, while the former is used by ingress TE routers, or Path Computation Engines (PCEs) when optimizing the placement of a flow based on the current network resource utilization.

For example, when steering traffic on a delay optimized path, the IGP can use its link state database's view of the network topology to compute a path optimizing for the delay metric of each link in the network resulting in a cumulative lowest delay path.

When path selection is network state dependent, the path computation can leverage Traffic Engineering mechanisms (e.g., as defined in [RFC2702]) to compute feasible paths taking into account the incoming traffic demand rate and current state of network. This allows avoiding overly utilized links, and reduces the chance of congestion on traversed links.

To enable TE path placement, the link state is advertised with current reservations, thereby reflecting the available bandwidth on each link. Such link reservations may be maintained centrally on a network wide network resource manager, or distributed on devices (as usually done with RSVP). TE extensions exist today to allow IGPs (e.g., [RFC3630] and [RFC5305]), and BGP-LS [RFC7752] to advertise such link state reservations.

When network resource reservations are also slice aggregate aware, the link state can carry per slice aggregate state (e.g., reservable bandwidth). This allows path computation to take into account the specific network resources available for a slice aggregate when determining the path for a specific flow. In this case, we refer to the process of path placement and path provisioning as slice aggregate aware TE.

4. Slice Policy Modes

A slice policy can be used to dictate if the partitioning of the shared network resources amongst multiple slice aggregates can be achieved by realizing slice aggregates in:

a) data plane only, or

b) control plane only, or

c) both control and data planes.

4.1. Data plane Slice Policy Mode

The physical network resources can be partitioned on network devices by applying a Per Hop forwarding Behavior (PHB) onto packets that traverse the network devices. In the Diffserv model, a Class Selector (CS) is carried in the packet and is used by transit nodes to apply the PHB that determines the scheduling treatment and drop probability for packets.

When data plane slice policy mode is applied, packets need to be forwarded on the specific slice aggregate network resources and need to be applied a specific forwarding treatment that is dictated in the slice policy (refer to Section 5.1 below). A Slice Selector (SS) MUST be carried in each packet to identify the slice aggregate that it belongs to.

The ingress node of a slice policy domain, in addition to marking packets with a Diffserv CS, MAY also add an SS to each slice aggregate packet. The transit nodes within a slice policy domain MAY use the SS to associate packets with a slice aggregate and to determine the Slice policy Per Hop Behavior (S-PHB) that is applied to the packet (refer to Section 5.1.3 for further details). The CS MAY be used to apply a Diffserv PHB on to the packet to allow differentiation of traffic treatment within the same slice aggregate.

When data plane only slice policy mode is used, routers may rely on a network state independent view of the topology to determine the best paths to reach destinations. In this case, the best path selection

dictates the forwarding path of packets to the destination. The SS field carried in each packet determines the specific S-PHB treatment along the selected path.

For example, the Segment-Routing Flexible Algorithm [I-D.ietf-lsr-flex-algo] may be deployed in a network to steer packets on the IGP computed lowest cumulative delay path. A slice policy may be used to allow links along the least latency path to share its data plane resources amongst multiple slice aggregates. In this case, the packets that are steered on a specific slice policy carry the SS field that enables routers (along with the Diffserv CS) to determine the S-PHB and enforce slice aggregate traffic streams.

4.2. Control Plane Slice Policy Mode

The physical network resources in the network can be logically partitioned by having a representation of network resources appear in a virtual topology. The virtual topology can contain all or a subset of the physical network resources by applying specific topology filters on the native topology. The logical network resources that appear in the virtual topology can reflect a part, whole, or in-excess of the physical network resource capacity (when oversubscription is desirable). For example, a physical link bandwidth can be divided into fractions, each dedicated to a slice aggregate. Each fraction of the physical link bandwidth MAY be represented as a logical link in a virtual topology that is used when determining paths associated with a specific slice aggregate. The virtual topology associated with the slice policy can be used by routing protocols, or by the ingress/PCE when computing slice aggregate aware TE paths.

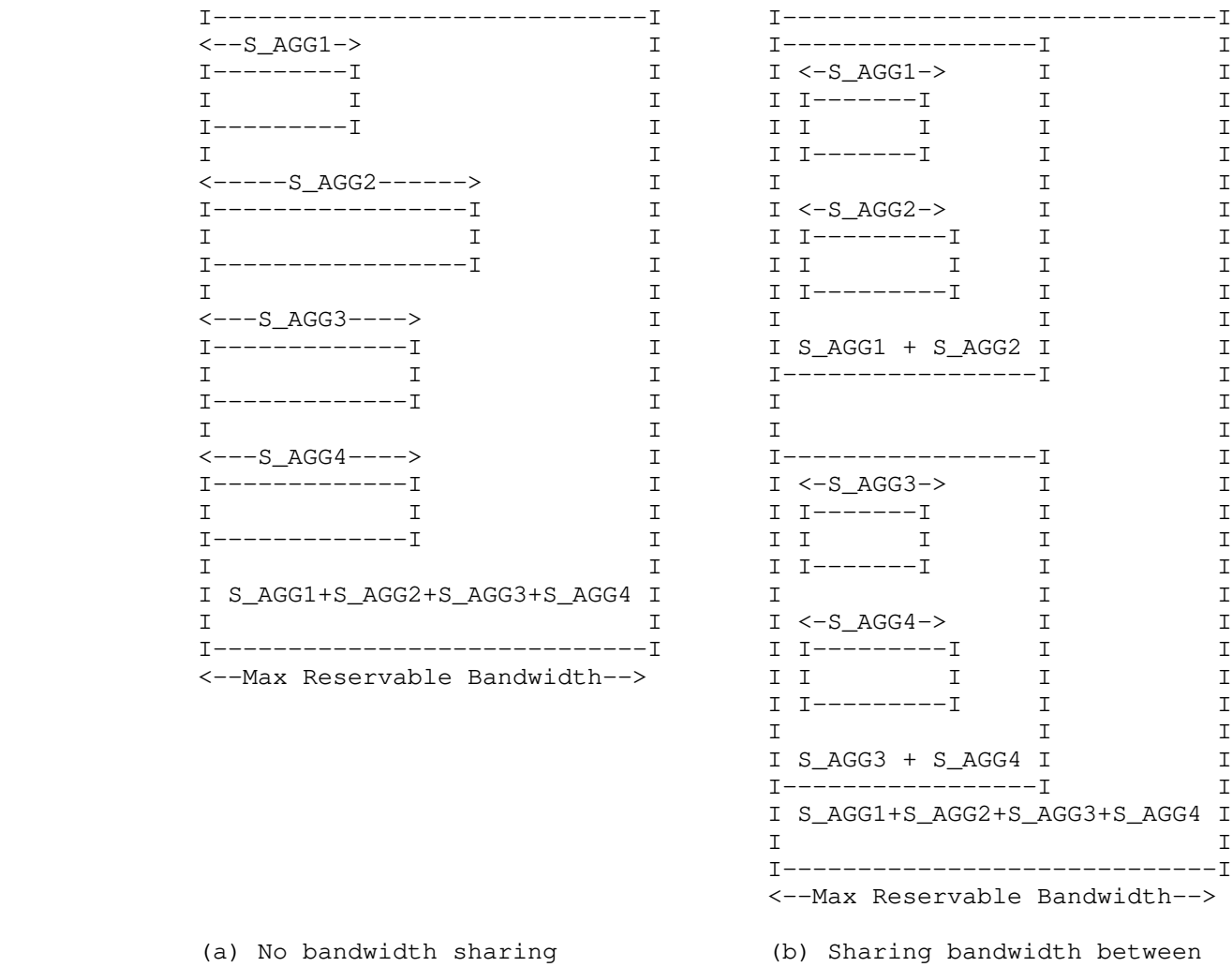
To perform network state dependent path computation in this mode (slice aggregate aware TE), the resource reservation on each link needs to be slice aggregate aware. Details of required IGP extensions to support SA-TE are described in [I-D.bestbar-lsr-slice-aware-te].

The same physical link may be member of multiple slice policies that instantiate different slice aggregates. The slice aggregate network resource availability on such a link is updated (and may be advertised) whenever new paths are placed in the network. The slice aggregate resource reservation, in this case, MAY be maintained on each device or off the device on a resource reservation manager that holds reservation states for those links in the network.

Multiple slice aggregates can form a group and share the available network resources allocated to each slice aggregate. In this case, a node can update the reservable bandwidth for each slice aggregate to

take into consideration the available bandwidth from other slice aggregates in the same group.

For illustration purposes, the diagram below represents bandwidth isolation or sharing amongst a group of slice aggregates. In Figure 1a, the slice aggregates: S_AGG1, S_AGG2, S_AGG3 and S_AGG4 are not sharing any bandwidths between each other. In Figure 1b, the slice aggregates: S_AGG1 and S_AGG2 can share the available bandwidth portion allocated to each amongst them. Similarly, S_AGG3 and S_AGG4 can share amongst themselves any available bandwidth allocated to them, but they cannot share available bandwidth allocated to S_AGG1 or S_AGG2. In both cases, the Max Reservable Bandwidth may exceed the actual physical link resource capacity to allow for over subscription.



(a) No bandwidth sharing between slice aggregates.

(b) Sharing bandwidth between slice aggregates of the same group

Figure 1: Bandwidth Isolation/Sharing.

4.3. Data and Control Plane Slice Policy Mode

In order to support strict guarantees for slice aggregates, the network resources can be partitioned in both the control plane and data plane.

The control plane partitioning allows the creation of customized topologies per slice aggregate that routers or a Path Computation

Engine (PCE) can use to determine optimal path placement for specific demand flows (Slice aggregate aware TE).

The data plane partitioning protects slice aggregate traffic from network resource contention that could occur due to bursts in traffic from other slice aggregates traversing the same shared network resource.

5. Slice Policy Instantiation

A network slice can span multiple technologies and multiple administrative domains. Depending on the network slice consumer's requirements, a network slice can be differentiated from other network slices in terms of data, control or management planes.

The consumer of a network slice expresses their intent by specifying requirements rather than mechanisms to realize the slice. The requirements for a network slice can vary and can be expressed in terms of connectivity needs between end-points (point-to-point, point-to-multipoint or multipoint-to-multipoint) with customizable network capabilities that may include data speed, quality, latency, reliability, security, and services (refer to [I-D.ietf-teas-ietf-network-slice-definition] for more details). These capabilities are always provided based on a Service Level Agreement (SLA) between the network slice consumer and the provider.

The onus is on the network slice controller to consume the service layer slice intent and realize it with an appropriate slice policy. Multiple IETF network slices can be mapped to the same slice policy resulting in a slice aggregate. The network wide consistent slice policy definition is distributed to the devices in the network as shown in Figure 2. The specification of the network slice intent on the northbound interface of the controller and the mechanism used to map the network slice to a slice policy are outside the scope of this document.

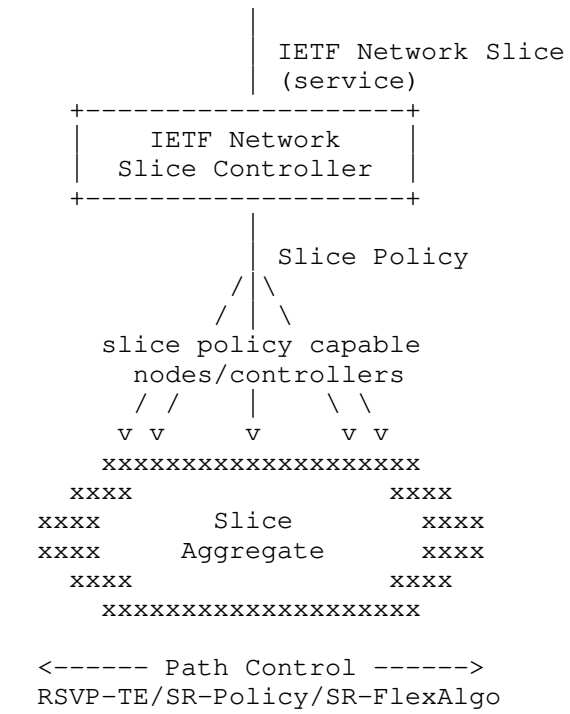


Figure 2: Slice Policy Instantiation.

5.1. Slice Policy Definition

The slice policy is network-wide construct that is consumed by network devices, and may include rules that control the following:

- Data plane specific policies: This includes the SS, any firewall rules or flow-spec filters, and QoS profiles associated with the slice policy and any classes within it.

- Control plane specific policies: This includes guaranteed bandwidth, any network resource sharing amongst slice policies, and reservation preference to prioritize any reservations of a specific slice policy over others.

- Topology membership policies: This defines topology filter policies that dictate node/link/function network resource topology association for a specific slice policy.

There is a desire for flexibility in realizing network slices to support the services across networks consisting of products from multiple vendors. These networks may also be grouped into disparate

domains and deploy various path control technologies and tunnel techniques to carry traffic across the network. It is expected that a standardized data model for slice policy will facilitate the instantiation and management of slice aggregates on slice policy capable nodes. A YANG data model for the slice policy instantiation on network devices is described in [I-D.bestbar-teas-yang-slice-policy].

It is also possible to distribute the slice policy to network devices using several mechanisms, including protocols such as NETCONF or RESTCONF, or exchanging it using a suitable routing protocol that network devices participate in (such as IGP(s) or BGP). The extensions to enable specific protocols to carry a slice policy definition will be described in separate documents.

5.1.1. Slice Policy Data Plane Selector

A router MUST be able to identify a packet belonging to a slice aggregate before it can apply the associated forwarding treatment or S-PHB. One or more fields within the packet MAY be used as an SS to do this.

Forwarding Address Based Slice Selector:

It is possible to assign a different forwarding address (or MPLS forwarding label in case of MPLS network) for each slice aggregate on a specific node in the network. [RFC3031] states in Section 2.1 that: 'Some routers analyze a packet's network layer header not merely to choose the packet's next hop, but also to determine a packet's "precedence" or "class of service"'. Assigning a unique forwarding address (or MPLS forwarding label) to each slice aggregate allows slice aggregate packets destined to a node to be distinguished by the destination address (or MPLS forwarding label) that is carried in the packet.

This approach requires maintaining per slice aggregate state for each destination in the network in both the control and data plane and on each router in the network. For example, consider a network slicing provider with a network composed of 'N' nodes, each with 'K' adjacencies to its neighbors. Assuming a node can be reached over 'M' different slice aggregates, the node assigns and advertises reachability to 'N' unique forwarding addresses, or MPLS forwarding labels. Similarly, each node assigns a unique forwarding address (or MPLS forwarding label) for each of its 'K' adjacencies to enable strict steering over the adjacency for each slice. The total number of control and data plane states that need to be stored and programmed in a router's forwarding is (N+K)*M states. Hence, as 'N', 'K', and 'M' parameters increase,

this approach suffers from scalability challenges in both the control and data planes.

Global Identifier Based Slice Selector:

A slice policy MAY include a Global Identifier Slice Selector (GISS) field as defined in [I-D.kompella-mpls-mspl4fa] that is carried in each packet in order to associate it to a specific slice aggregate, independent of the forwarding address or MPLS forwarding label that is bound to the destination. Routers within the slice policy domain can use the forwarding address (or MPLS forwarding label) to determine the forwarding next-hop(s), and use the GISS field in the packet to infer the specific forwarding treatment that needs to be applied on the packet.

The GISS can be carried in one of multiple fields within the packet, depending on the dataplane used. For example, in MPLS networks, the GISS can be encoded within an MPLS label that is carried in the packet's MPLS label stack. All packets that belong to the same slice aggregate MAY carry the same GISS in the MPLS label stack. It is also possible to have multiple GISS's map to the same slice aggregate.

The GISS can be encoded in an MPLS label and may appear in several positions in the MPLS label stack. For example, the VPN service label may act as a GISS to allow VPN packets to be associated with a specific slice aggregate. In this case, a single VPN service label acting as a GISS MAY be allocated by all Egress PEs of a VPN. Alternatively, multiple VPN service labels MAY act as GISS's that map a single VPN to the same slice aggregate to allow for multiple Egress PEs to allocate different VPN service labels for a VPN. In other cases, a range of VPN service labels acting as multiple GISS's MAY map multiple VPN traffic to a single slice aggregate. An example of such deployment is shown in Figure 3.

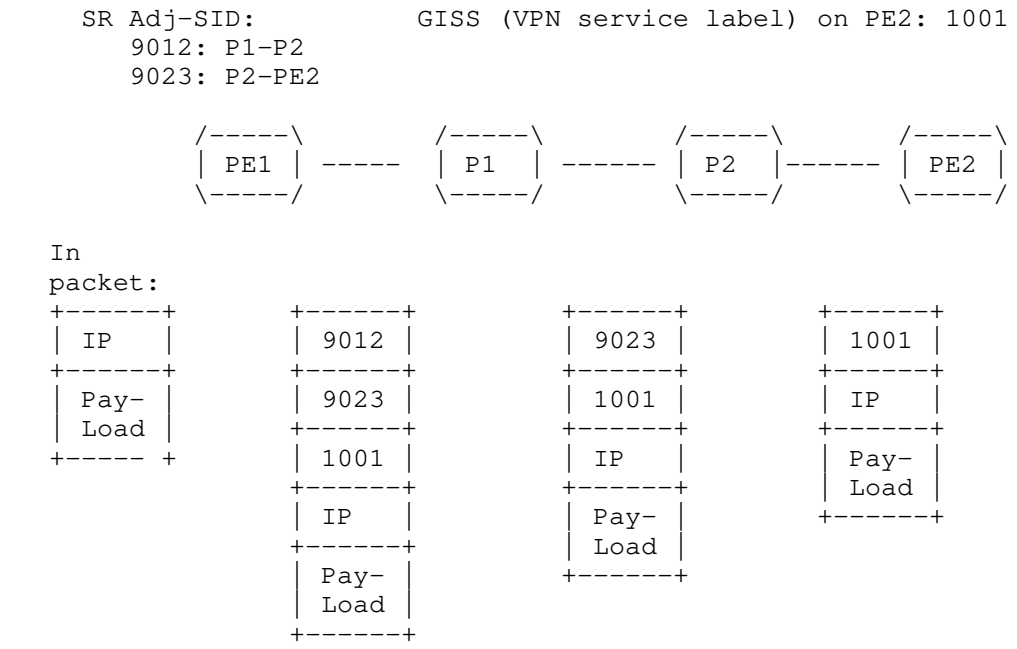


Figure 3: GISS or VPN label at bottom of label stack.

In some cases, the position of the GISS may not be at a fixed position in the MPLS label header. In this case, the GISS label can show up in any position in the MPLS label stack. To enable a transit router to identify the position of the GISS label, a special purpose label (ideally a base special purpose label (bSPL)) can be used as a GISS label indicator. [I-D.kompella-mpls-mspl4fa] proposes a new bSPL called Forwarding Actions Identifier (FAI) that is assigned to alert of the presence of multiple actions and action data (including the presence of the GISS) that are carried within the MPLS label stack. The slice policy ingress boundary node, in this case, imposes two labels: the FAI label and a forwarding actions label that includes the GISS to identify the slice aggregate that packets belong to as shown in Figure 4.

[I-D.decraene-mpls-slid-encoded-entropy-label-id] also proposes to repurpose the ELI/EL [RFC6790] to carry the Slice Identifier in order to minimize the size of the MPLS stack and ease incremental deployment.

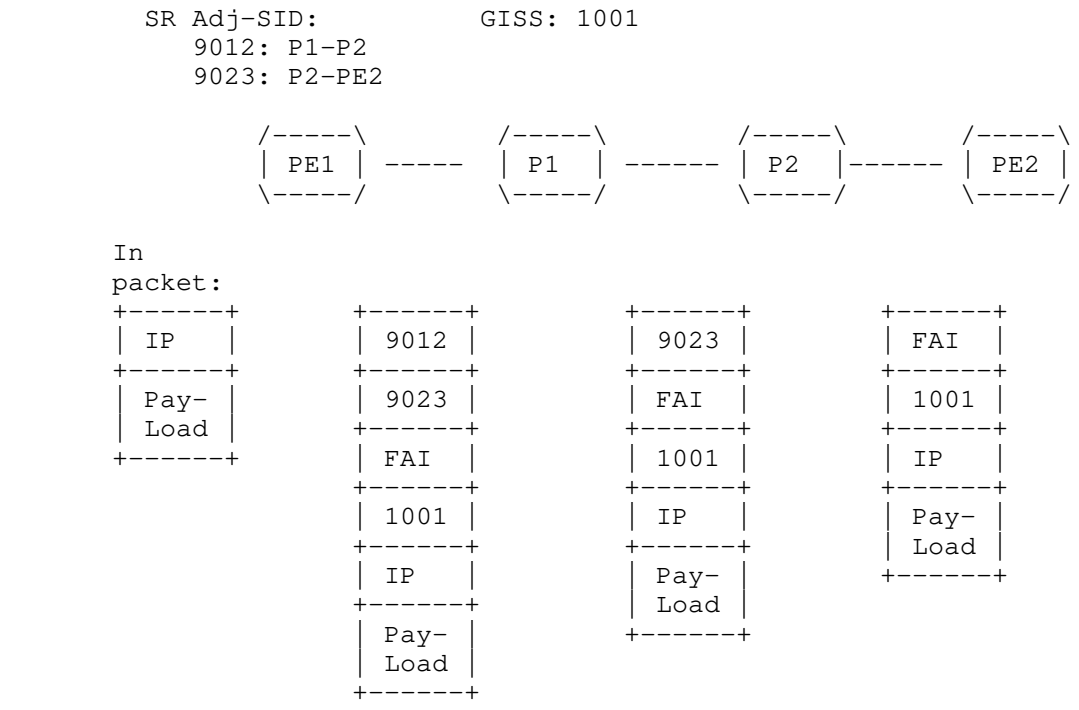


Figure 4: FAI and GISS label in the label stack.

When the slice is realized over an IP dataplane, the GISS can be encoded in the IP header. For example, the SSL can be encoded in portion of the IPv6 Flow Label field as described in [I-D.filsfils-spring-srv6-stateless-slice-id].

5.1.2. Slice Policy Resource Reservation

Bandwidth and network resource allocation strategies for slice policies are essential to achieve optimal placement of paths within the network while still meeting the target SLOs.

Resource reservation allows for the managing of available bandwidth and for prioritization of existing allocations to enable preference-based preemption when contention on a specific network resource arises. Sharing of a network resource's available bandwidth amongst a group of slice policies may also be desirable. For example, a slice aggregate may not always be using all of its reservable bandwidth; this allows other slice policies in the same group to use the available bandwidth resources.

Congestion on shared network resources may result from sub-optimal placement of paths in different slice policies. When this occurs, preemption of some slice aggregate specific paths may be desirable to alleviate congestion. A preference based allocation scheme enables prioritization of slice aggregate paths that can be preempted.

Since network characteristics and its state can change over time, the slice policy topology and its state also need to be propagated in the network to enable ingress TE routers or Path Computation Engine (PCEs) to perform accurate path placement based on the current state of the slice policy network resources.

5.1.3. Slice Policy Per Hop Behavior

In Diffserv terminology, the forwarding behavior that is assigned to a specific class is called a Per Hop Behavior (PHB). The PHB defines the forwarding precedence that a marked packet with a specific CS receives in relation to other traffic on the Diffserv-aware network.

A Slice policy Per Hop Behavior (S-PHB) is the externally observable forwarding behavior applied to a specific packet belonging to a slice aggregate. The goal of an S-PHB is to provide a specified amount of network resources for traffic belonging to a specific slice aggregate. A single slice policy may also support multiple forwarding treatments or services that can be carried over the same logical network.

The slice aggregate traffic may be identified at slice policy ingress boundary nodes by carrying a SS to allow routers to apply a specific forwarding treatment that guarantee the SLA(s).

With Differentiated Services (Diffserv) it is possible to carry multiple services over a single converged network. Packets requiring the same forwarding treatment are marked with a Class Selector (CS) at domain ingress nodes. Up to eight classes or Behavior Aggregates (BAs) may be supported for a given Forwarding Equivalence Class (FEC) [RFC2475]. To support multiple forwarding treatments over the same slice aggregate, a slice aggregate packet MAY also carry a Diffserv CS to identify the specific Diffserv forwarding treatment to be applied on the traffic belonging to the same slice policy.

At transit nodes, the CS field carried inside the packets are used to determine the specific PHB that determines the forwarding and scheduling treatment before packets are forwarded, and in some cases, drop probability for each packet.

5.1.4. Slice Policy Topology

A key element of the slice policy is a customized topology that may include the full or subset of the physical network topology. The slice policy topology could also span multiple administrative domains and/or multiple dataplane technologies.

A slice policy topology can overlap or share a subset of links with another slice policy topology. A number of topology filtering policies can be defined as part of the slice policy to limit the specific topology elements that belong to a slice policy. For example, a topology filtering policy can leverage Resource Affinities as defined in [RFC2702] to include or exclude certain links for a specific slice aggregate. The slice policy may also include a reference to a predefined topology (e.g., derived from a Flexible Algorithm Definition (FAD) as defined in [I-D.ietf-lsr-flex-algo], or Multi-Topology ID as defined [RFC4915].

5.2. Slice Policy Boundary

A network slice originates at the edge nodes of a network slice provider. Traffic that is steered over the corresponding slice aggregate may traverse slice policy capable interior nodes as well as slice policy incapable interior nodes.

The network slice may encompass one or more domains administered by a provider. For example, an organization's intranet or an ISP. The network provider is responsible for ensuring that adequate network resources are provisioned and/or reserved to support the SLAs offered by the network end-to-end.

5.2.1. Slice Policy Edge Nodes

Slice policy edge nodes sit at the boundary of a network slice provider network and receive traffic that requires steering over network resources specific to a slice aggregate. These edge nodes are responsible for identifying slice aggregate specific traffic flows by possibly inspecting multiple fields from inbound packets (e.g., implementations may inspect IP traffic's network 5-tuple in the IP and transport protocol headers) to decide on which slice policy it can be steered.

Network slice ingress nodes may condition the inbound traffic at network boundaries in accordance with the requirements or rules of each service's SLAs. The requirements and rules for network slice services are set using mechanisms which are outside the scope of this document.

When data plane slice policy is applied, the slice policy ingress boundary nodes are responsible for adding a suitable SS onto packets that belong to specific slice aggregate. In addition, edge nodes MAY mark the corresponding Diffserv CS to differentiate between different types of traffic carried over the same slice aggregate.

5.2.2. Slice Policy Interior Nodes

A slice policy interior node receives slice traffic and MAY be able to identify the packets belonging to a specific slice aggregate by inspecting the SS field carried inside each packet, or by inspecting other fields within the packet that may identify the traffic streams that belong to a specific slice aggregate. For example, when data plane slice policy is applied, interior nodes can use the SS carried within the packet to apply the corresponding S-PHB forwarding behavior. Nodes within the network slice provider network may also inspect the Diffserv CS within each packet to apply a per Diffserv class PHB within the slice policy, and allow differentiation of forwarding treatments for packets forwarded over the same slice aggregate network resources.

5.2.3. Slice Policy Incapable Nodes

Packets that belong to a slice aggregate may need to traverse nodes that are slice policy incapable. In this case, several options are possible to allow the slice traffic to continue to be forwarded over such devices and be able to resume the slice policy forwarding treatment once the traffic reaches devices that are slice policy capable.

When data plane slice policy is applied, packets carry a SS to allow slice interior nodes to identify them. To enable end-to-end network slicing, the SS MUST be maintained in the packets as they traverse devices within the network - including slice policy incapable devices.

For example, when the SS is an MPLS label at the bottom of the MPLS label stack, packets can traverse over devices that are slice policy incapable without any further considerations. On the other hand, when the SSL is at the top of the MPLS label stack, packets can be bypassed (or tunneled) over the slice policy incapable devices towards the next device that supports slice policy as shown in Figure 5.

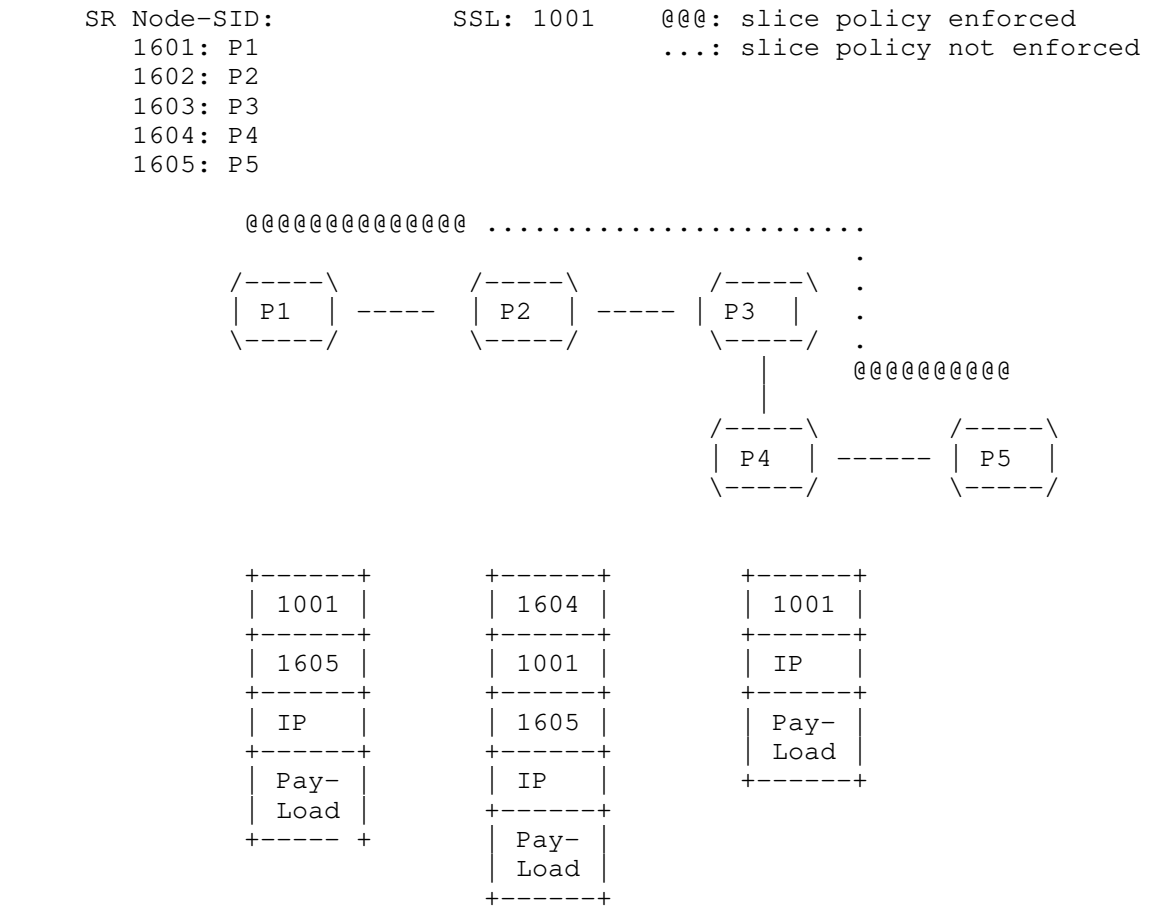


Figure 5: Extending network slice over slice policy incapable device(s).

5.2.4. Combining Slice Policy Modes

It is possible to employ a combination of the slice policy modes that were discussed in Section 4 to realize a network slice. For example, data and control plane slice policy mode can be employed in parts of a network, while control plane slice policy mode can be employed in the other parts of the network. The path selection, in such case, can take into account the slice aggregate specific available network resources. The SS carried within packets allow transit nodes to enforce the corresponding S-PHB on the parts of the network that apply the data plane slice policy mode. The SS can be maintained while traffic traverses nodes that do not enforce data plane slice

policy mode, and so slice PHB enforcement can resume once traffic traverses capable nodes.

5.3. Mapping Traffic on Slice Aggregates

The usual techniques to steer traffic onto paths can be applicable when steering traffic over paths established for a specific slice aggregate.

For example, one or more (layer-2 or layer-3) VPN services can be directly mapped to paths established for a slice aggregate. In this case, the per Virtual Routing and Forwarding (VRF) instance traffic that arrives on the Provider Edge (PE) router over external interfaces can be directly mapped to a specific slice aggregate path. External interfaces can be further partitioned (e.g., using VLANs) to allow mapping one or more VLANs to specific slice aggregate paths.

Another option is steer traffic to specific destinations directly over multiple slice policies. This allows traffic arriving on any external interface and targeted to such destinations to be directly steered over the slice paths.

A third option that can also be used is to utilize a data plane firewall filter or classifier to enable matching of several fields in the incoming packets to decide whether the packet is steered on a specific slice aggregate. This option allows for applying a rich set of rules to identify specific packets to be mapped to a slice aggregate. However, it requires data plane network resources to be able to perform the additional checks in hardware.

6. Control Plane Extensions

Routing protocols may need to be extended to carry additional per slice aggregate link state. For example, [RFC5305], [RFC3630], and [RFC7752] are ISIS, OSPF, and BGP protocol extensions to exchange network link state information to allow ingress TE routers and PCE(s) to do proper path placement in the network. The extensions required to support network slicing may be defined in other documents, and are outside the scope of this document.

The instantiation of a slice policy may need to be automated. Multiple options are possible to facilitate automation of distribution of a slice policy to capable devices.

For example, a YANG data model for the slice policy may be supported on network devices and controllers. A suitable transport (e.g., NETCONF [RFC6241], RESTCONF [RFC8040], or gRPC) may be used to enable configuration and retrieval of state information for slice policies

on network devices. The slice policy YANG data model is outside the scope of this document, and is defined in [I-D.bestbar-teas-yang-slice-policy].

7. Applicability to Path Control Technologies

The slice policy modes described in this document are agnostic to the technology used to setup paths that carry slice aggregate traffic. One or more paths connecting the endpoints of the mapped IETF network slices may be selected to steer the corresponding traffic streams over the resources allocated for the slice aggregate.

For example, once the feasible paths within a slice policy topology are selected, it is possible to use RSVP-TE protocol [RFC3209] to setup or signal the LSPs that would be used to carry the slice aggregate traffic. Specific extensions to RSVP-TE protocol to enable signaling of slice aggregate aware RSVP LSPs are outside the scope of this document.

Alternatively, Segment Routing (SR) [RFC8402] may be used and the feasible paths can be realized by steering over specific segments or segment-lists using an SR policy. Further details on how the slice policy modes presented in this document can be realized over an SR network is discussed in [I-D.bestbar-spring-scalable-ns], and [I-D.bestbar-lsr-spring-sa].

8. IANA Considerations

This document has no IANA actions.

9. Security Considerations

The main goal of network slicing is to allow for varying treatment of traffic from multiple different network slices that are utilizing a common network infrastructure and to allow for different levels of services to be provided for traffic traversing a given network resource.

A variety of techniques may be used to achieve this, but the end result will be that some packets may be mapped to specific resources and may receive different (e.g., better) service treatment than others. The mapping of network traffic to a specific slice policy is indicated primarily by the SS, and hence an adversary may be able to utilize resources allocated to a specific slice policy by injecting packets carrying the same SS field in their packets.

Such theft-of-service may become a denial-of-service attack when the modified or injected traffic depletes the resources available to forward legitimate traffic belonging to a specific slice policy.

The defense against this type of theft and denial-of-service attacks consists of a combination of traffic conditioning at slice policy domain boundaries with security and integrity of the network infrastructure within a slice policy domain.

10. Acknowledgement

The authors would like to thank Krzysztof Szarkowicz, Swamy SRK, Navaneetha Krishnan, Prabhu Raj Villadathu Karunakaran and Jie Dong for their review of this document, and for providing valuable feedback on it.

11. Contributors

The following individuals contributed to this document:

Colby Barth
Juniper Networks
Email: cbarth@juniper.net

Srihari R. Sangli
Juniper Networks
Email: ssangli@juniper.net

Chandra Ramachandran
Juniper Networks
Email: csekar@juniper.net

12. References

12.1. Normative References

[I-D.bestbar-lsr-slice-aware-te]
Britto, W., Shetty, R., Barth, C., Wen, B., Peng, S., and R. Chen, "IGP Extensions for Support of Slice Aggregate Aware Traffic Engineering", draft-bestbar-lsr-slice-aware-te-00 (work in progress), February 2021.

[I-D.bestbar-lsr-spring-sa]
Saad, T., Beeram, V. P., Chen, R., Peng, S., Wen, B., and D. Ceccarelli, "IGP Extensions for SR Slice Aggregate SIDs", draft-bestbar-lsr-spring-sa-00 (work in progress), February 2021.

[I-D.bestbar-spring-scalable-ns]
Saad, T., Beeram, V. P., Chen, R., Peng, S., Wen, B., and D. Ceccarelli, "Scalable Network Slicing over SR Networks", draft-bestbar-spring-scalable-ns-01 (work in progress), February 2021.

[I-D.bestbar-teas-yang-slice-policy]
Saad, T., Beeram, V. P., Wen, B., Ceccarelli, D., Peng, S., Chen, R., Contreras, L. M., and X. Liu, "YANG Data Model for Slice Policy", draft-bestbar-teas-yang-slice-policy-00 (work in progress), February 2021.

[I-D.decraene-mpls-slid-encoded-entropy-label-id]
Decraene, B., Filsfils, C., Henderickx, W., Saad, T., Beeram, V. P., and L. Jalil, "Using Entropy Label for Network Slice Identification in MPLS networks.", draft-decraene-mpls-slid-encoded-entropy-label-id-01 (work in progress), February 2021.

[I-D.filsfils-spring-srv6-stateless-slice-id]
Filsfils, C., Clad, F., Camarillo, P., and K. Raza, "Stateless and Scalable Network Slice Identification for SRv6", draft-filsfils-spring-srv6-stateless-slice-id-02 (work in progress), January 2021.

[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-15 (work in progress), April 2021.

[I-D.kompella-mpls-mspl4fa]
Kompella, K., Beeram, V. P., Saad, T., and I. Meilik, "Multi-purpose Special Purpose Label for Forwarding Actions", draft-kompella-mpls-mspl4fa-00 (work in progress), February 2021.

[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>.

[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>.

[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>.

[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>.

[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>.

[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>.

12.2. Informative References

[I-D.ietf-teas-ietf-network-slice-definition] Rokui, R., Homma, S., Makhijani, K., Contreras, L. M., and J. Tantsura, "Definition of IETF Network Slices", draft-ietf-teas-ietf-network-slice-definition-01 (work in progress), February 2021.

[I-D.ietf-teas-rfc3272bis]
Farrel, A., "Overview and Principles of Internet Traffic Engineering", draft-ietf-teas-rfc3272bis-11 (work in progress), April 2021.

[I-D.nsdt-teas-ns-framework]
Gray, E. and J. Drake, "Framework for IETF Network Slices", draft-nsdt-teas-ns-framework-05 (work in progress), February 2021.

[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>.

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, DOI 10.17487/RFC2702, September 1999, <https://www.rfc-editor.org/info/rfc2702>.

[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>.

[RFC8040] Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", RFC 8040, DOI 10.17487/RFC8040, January 2017, <https://www.rfc-editor.org/info/rfc8040>.

Authors' Addresses

Tarek Saad
Juniper Networks

Email: tsaad@juniper.net

Vishnu Pavan Beeram
Juniper Networks

Email: vbeeram@juniper.net

Bin Wen
Comcast

Email: Bin_Wen@cable.comcast.com

Daniele Ceccarelli
Ericsson

Email: daniele.ceccarelli@ericsson.com

Joel Halpern
Ericsson

Email: joel.halpern@ericsson.com

Shaofu Peng
ZTE Corporation

Email: peng.shaofu@zte.com.cn

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.cn

Xufeng Liu
Volta Networks

Email: xufeng.liu.ietf@gmail.com

Luis M. Contreras
Telefonica

Email: luismiguel.contrerasmurillo@telefonica.com

Reza Rokui
Nokia

Email: reza.rokui@nokia.com