Workgroup: TEAS Working Group
Internet-Draft: draft-ietf-teas-ns-ip-mpls-00
Published: 16 June 2022
Intended Status: Informational
Expires: 18 December 2022
Authors: T. Saad
Juniper Networks
V. Beeram
Juniper Networks
J. Dong
Huawei Technologies
B. Wen
Comcast
D. Ceccarelli
Ericsson
J. Halpern
Ericsson
S. Peng
ZTE Corporation
R. Chen
ZTE Corporation
X. Liu
Volta Networks
L. Contreras
Telefonica
R. Rokui
Ciena
L. Jalil
Verizon

Realizing Network Slices in IP/MPLS Networks

Abstract

Realizing network slices may require the Service Provider to have 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. Multiple network slices can be realized on the same network while ensuring slice elasticity in terms of network resource allocation. This document describes a scalable solution to realize network slicing in IP/MPLS networks by supporting multiple services on top of a single physical network by relying on compliant domains and nodes to provide forwarding treatment (scheduling, drop policy, resource usage) on to packets that carry identifiers that indicate the slicing service that is to be applied to the packets.

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 18 December 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction
  1.1. Terminology
  1.2. Acronyms and Abbreviations
2. Network Resource Slicing Membership
3. IETF Network Slice Realization
  3.1. Network Topology Filters
  3.2. IETF Network Slice Service Request
  3.3. Slice-Flow Aggregation
  3.4. Path Placement over NRP Filter Topology
  3.5. NRP Policy Installation
  3.6. Path Instantiation
  3.7. Service Mapping
4. Network Resource Partition Modes
  4.1. Data plane Network Resource Partition Mode
  4.2. Control Plane Network Resource Partition Mode
  4.3. Data and Control Plane Network Resource Partition Mode
5. Network Resource Partition Instantiation
  5.1. NRP Policy Definition
    5.1.1. Network Resource Partition - Flow-Aggregate Selector
    5.1.2. Network Resource Partition Resource Reservation
    5.1.3. Network Resource Partition Per Hop Behavior
    5.1.4. Network Resource Partition Topology
  5.2. Network Resource Partition Boundary
    5.2.1. Network Resource Partition Edge Nodes
    5.2.2. Network Resource Partition Interior Nodes
    5.2.3. Network Resource Partition Incapable Nodes
    5.2.4. Combining Network Resource Partition Modes
6. Mapping Traffic on Slice-Flow Aggregates
  6.1. Network Slice-Flow Aggregate Relationships
7. Path Selection and Instantiation
  7.1. Applicability of Path Selection to Slice-Flow Aggregates

7.2. Applicability of Path Control Technologies to Slice-Flow Aggregates
7.2.1. RSVP-TE Based Slice-Flow Aggregate Paths
7.2.2. SR Based Slice-Flow Aggregate Paths
8. Network Resource Partition Protocol Extensions
9. Outstanding Issues
10. IANA Considerations
11. Security Considerations
12. Acknowledgement
13. Contributors
14. References
14.1. Normative References
14.2. Informative References
Authors' Addresses

1. Introduction

Network slicing allows a Service Provider to create independent and logical networks on top of a shared physical network infrastructure. Such network slices can be offered to customers or used internally by the Service Provider to enhance the delivery of their service offerings. A Service Provider can also use network slicing to structure and organize the elements of its infrastructure. The solution discussed in this document works with any path control technology (such as RSVP-TE, or SR) that can be used by a Service Provider to realize network slicing in IP/MPLS networks.

[I-D.ietf-teas-ietf-network-slices] provides the definition of a network slice for use within the IETF and discusses the general framework for requesting and operating IETF Network Slices, their characteristics, and the necessary system components and interfaces. It also discusses 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-Flow Aggregate which comprises of one or more IETF network slice traffic streams. It also describes the Network Resource Partition (NRP) and the NRP Policy that can be used to instantiate control and data plane behaviors on select topological elements associated with the NRP that supports a Slice-Flow Aggregate - refer Section 5.1 for further details.

The IETF Network Slice Controller is responsible for the aggregation of multiple IETF network traffic streams into a Slice-Flow Aggregate, and for maintaining the mapping required between them. The mechanisms used by the controller to determine the mapping of one or more IETF network slice to a Slice-Flow Aggregate 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 Diffserv (DS) domain [RFC2475], packets requiring the same forwarding treatment (scheduling and drop policy) are classified and marked with the respective Class Selector (CS) Codepoint (or the Traffic Class (TC) field for MPLS packets [RFC5462]) at the DS domain ingress nodes. Such packets are said to belong to a Behavior Aggregate (BA) that has a common set of behavioral characteristics or a common set of delivery requirements. At transit nodes, the CS is inspected to determine the specific forwarding treatment to be applied before the packet is forwarded. A similar approach is adopted in this document to realize network slicing. The solution proposed in this document does not mandate Diffserv to be enabled in the network to provide a specific forwarding treatment.

When logical networks associated with an NRP are realized on top of a shared physical network infrastructure, it is important to steer traffic on the specific network resources partition that is allocated for a given Slice-Flow Aggregate. In packet networks, the packets of a specific Slice-Flow Aggregate may be identified by one or more specific fields carried within the packet. An NRP ingress boundary node (where Slice-Flow Aggregate traffic enters the NRP) populates the respective field(s) in packets that are mapped to a Slice-Flow Aggregate in order to allow interior NRP nodes to identify and apply the specific Per NRP Hop Behavior (NRP-PHB) associated with the Slice-Flow Aggregate. The NRP-PHB defines the scheduling treatment and, in some cases, the packet drop probability.

If Diffserv is enabled within the network, the Slice-Flow Aggregate traffic can further carry a Diffserv CS to enable differentiation of forwarding treatments for packets within a Slice-Flow Aggregate.

For example, when using MPLS as a dataplane, it is possible to identify packets belonging to the same Slice-Flow Aggregate by carrying an identifier in an MPLS Label Stack Entry (LSE). 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 NRP.

This document covers different modes of NRPs and discusses how each mode can ensure proper placement of Slice-Flow Aggregate paths and respective treatment of Slice-Flow Aggregate traffic.

1.1. Terminology

The reader is expected to be familiar with the terminology specified in [I-D.ietf-teas-ietf-network-slices].

The following terminology is used in the document:

IETF Network Slice:
refer to the definition of 'IETF network slice' in [I-D.ietf-teas-ietf-network-slices].

IETF Network Slice Controller (NSC):
refer to the definition in [I-D.ietf-teas-ietf-network-slices].

Network Resource Partition:
refer to the definition in [I-D.ietf-teas-ietf-network-slices].

Slice-Flow Aggregate:
a collection of packets that match an NRP Policy and are given the same forwarding treatment; a Slice-Flow Aggregate comprises of one or more IETF network slice traffic streams; the mapping of one or more IETF network slices to a Slice-Flow Aggregate is maintained by the IETF Network Slice Controller. The boundary nodes MAY also maintain a mapping of specific IETF network slice service(s) to a SFA.

Network Resource Partition Policy (NRP):
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 an NRP Policy results in the creation of an NRP.

NRP Identifier (NRP-ID):
an identifier that is globally unique within an NRP domain and that can be used in the control or management plane to identify the resources associated with the NRP.

NRP Capable Node:
a node that supports one of the NRP modes described in this document.

NRP Incapable Node:
a node that does not support any of the NRP modes described in this document.

Slice-Flow Aggregate Path:
a path that is setup over the NRP that is associated with a specific Slice-Flow Aggregate.

Slice-Flow Aggregate Packet:
a packet that traverses over the NRP that is associated with a specific Slice-Flow Aggregate.

NRP Filter Topology:
a set of topological elements associated with a Network Resource Partition.

NRP state aware TE (NRP-TE):
a mechanism for TE path selection that takes into account the available network resources associated with a specific NRP.

1.2. Acronyms and Abbreviations

BA: Behavior Aggregate

CS: Class Selector

NRP-PHB: NRP Per Hop Behavior as described in Section 5.1.3

FAS: Flow Aggregate Selector

FASL: Flow Aggregate Selector Label as described in Section 5.1.1

SLA: Service Level Agreements

SLO: Service Level Objectives

SLE: Service Level Expectations

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

AC: Attachment Circuit

CE: Customer Edge

PE: Provider Edge

PCEP: Path Computation Element (PCE) Communication Protocol (PCEP)

2. Network Resource Slicing Membership

An NRP that supports a Slice-Flow 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. The NRP topology may be

comprised of dedicated and/or shared network resources (e.g., in terms of processing power, storage, and bandwidth).

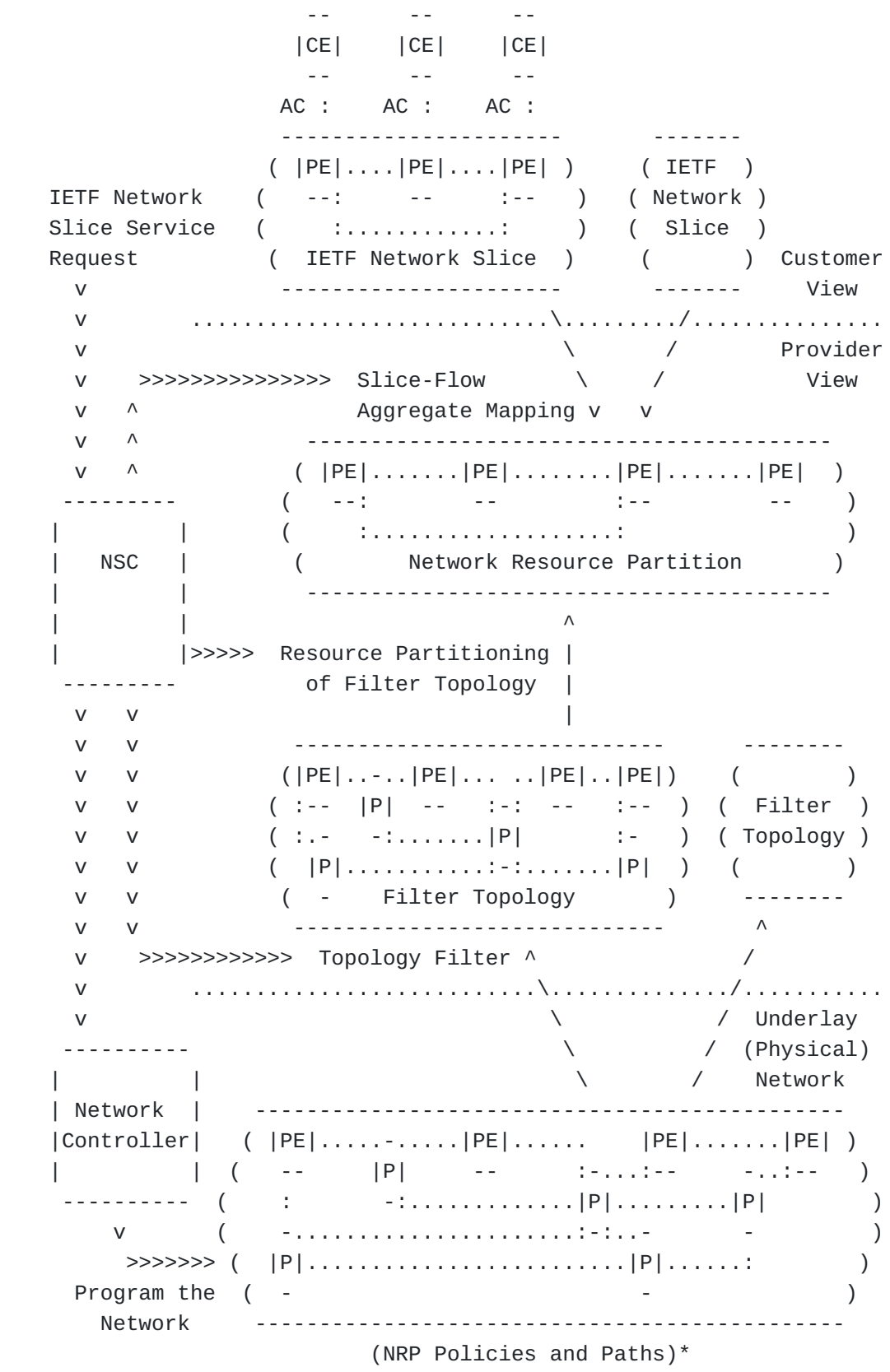
The physical network resources may be fully dedicated to a specific Slice-Flow Aggregate. For example, traffic belonging to a Slice-Flow Aggregate can traverse dedicated network resources without being subjected to contention from traffic of other Slice-Flow Aggregates. Dedicated physical network resource slicing allows for simple partitioning of the physical network resources amongst Slice-Flow Aggregates without the need to distinguish packets traversing the dedicated network resources since only one Slice-Flow Aggregate traffic stream can traverse the dedicated resource at any time.

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 NRPs that support multiple Slice-Flow Aggregates. The shared physical 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. IETF Network Slice Realization

Figure 1 describes the steps required to realize an IETF network slice service in a provider network using the solution proposed in this document. While Figure 4 of [I-D.ietf-teas-ietf-network-slices] provides an abstract architecture of an IETF Network Slice, this section intends to offer a realization of that architecture specific for IP/MPLS packet networks.

Each of the steps is further elaborated on in a subsequent section.



* : NRP Policy installation and path placement can be centralized or distributed.

Figure 1: IETF network slice realization steps.

3.1. Network Topology Filters

The Physical Network may be filtered into a number of Filter Topologies. Filter actions may include selection of specific nodes and links according to their capabilities and are based on network-wide policies. The resulting topologies can be used to host IETF Network Slices and provide a useful way for the network operator to know that all of the resources they are using to plan a network slice meet specific SLOs. This step can be done offline during planning activity, or could be performed dynamically as new demands arise.

Section 5.1.4 describes how topology filters can be associated with the NRP instantiated by the NRP Policy.

3.2. IETF Network Slice Service Request

The customer requests an IETF Network Slice Service specifying the CE-AC-PE points of attachment, the connectivity matrix, and the SLOs/SLEs as described in [I-D.ietf-teas-ietf-network-slices]. These capabilities are always provided based on a Service Level Agreement (SLA) between the network slice costumer and the provider.

This defines the traffic flows that need to be supported when the slice is realized. Depending on the mechanism and encoding of the Attachment Circuit (AC), the IETF Network Slice Service may also include information that will allow the operator's controllers to configure the PEs to determine what customer traffic is intended for this IETF Network Slice.

IETF Network Slice Service Requests are likely to arrive at various times in the life of the network, and may also be modified.

3.3. Slice-Flow Aggregation

A network may be called upon to support very many IETF Network Slices, and this could present scaling challenges in the operation of the network. In order to overcome this, the IETF Network Slice streams may be aggregated into groups according to similar characteristics.

A Slice-Flow Aggregate is a construct that comprises the traffic flows of one or more IETF Network Slices. The mapping of IETF Network Slices into an Slice-Flow Aggregate is a matter of local operator policy is a function executed by the Controller. The Slice-Flow Aggregate may be preconfigured, created on demand, or modified dynamically.

3.4. Path Placement over NRP Filter Topology

Depending on the underlying network technology, the paths are selected in the network in order to best deliver the SLOs for the different services carried by the Slice-Flow Aggregate. The path placement function (carried on ingress node or by a controller) is performed on the Filter Topology that is selected to support the Slice-Flow Aggregate.

Note that this step may indicate the need to increase the capacity of the underlying Filter Topology or to create a new Filter Topology.

3.5. NRP Policy Installation

A Controller function programs the physical network with policies for handling the traffic flows belonging to the Slice-Flow Aggregate. These policies instruct underlying routers how to handle traffic for a specific Slice-Flow Aggregate: the routers correlate markers present in the packets that belong to the Slice-Flow Aggregate. The way in which the NRP Policy is installed in the routers and the way that the traffic is marked is implementation specific. The NRP Policy instantiation in the network is further described in Section 5.

3.6. Path Instantiation

Depending on the underlying network technology, a Controller function may install the forwarding state specific to the Slice-Flow Aggregate so that traffic is routed along paths derived in the Path Placement step described in Section 3.4. The way in which the paths are instantiated is implementation specific.

3.7. Service Mapping

The edge points can be configured to support the network slice service by mapping the customer traffic to Slice-Flow Aggregates, possibly using information supplied when the IETF network slice service was requested. The edge points may also be instructed to mark the packets so that the network routers will know which policies and routing instructions to apply. The steering of traffic onto Slice-Flow Aggregate paths is further described in Section 6.

4. Network Resource Partition Modes

An NRP Policy can be used to dictate if the network resource partitioning of the shared network resources among multiple Slice-Flow Aggregates can be achieved:

a) in data plane only,

b) in control plane only, or

c) in both control and data planes.

4.1. Data plane Network Resource Partition 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) codepoint 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 NRP mode is applied, packets need to be forwarded on the specific NRP that supports the Slice-Flow Aggregate to ensure the proper forwarding treatment dictated in the NRP Policy is applied (refer to Section 5.1 below). In this case, a Flow Aggregate Selector (FAS) must be carried in each packet to identify the Slice-Flow Aggregate that it belongs to.

The ingress node of an NRP domain adds a FAS field if one is not already present in each Slice-Flow Aggregate packet. In the data plane NRP mode, the transit nodes within an NRP domain use the FAS to associate packets with a Slice-Flow Aggregate and to determine the Network Resource Partition Per Hop Behavior (NRP-PHB) that is applied to the packet (refer to Section 5.1.3 for further details). The CS is used to apply a Diffserv PHB on to the packet to allow differentiation of traffic treatment within the same Slice-Flow Aggregate.

When data plane only NRP mode is used, routers may rely on a network state independent view of the topology to determine the best paths. In this case, the best path selection dictates the forwarding path of packets to the destination. The FAS field carried in each packet determines the specific NRP-PHB treatment along the selected path.

4.2. Control Plane Network Resource Partition Mode

Multiple NRPs can be realized over the same set of physical resources. Each NRP is identified by an identifier (NRP-ID) that is globally unique within the NRP domain. The NRP state reservations for each NRP can be maintained on the network element or on a controller.

The network reservation states for a specific partition can be represented in a topology that contains all or a subset of the physical network elements (nodes and links) and reflect the network state reservations in that NRP. The logical network resources that appear in the NRP topology can reflect a part, whole, or in-excess

of the physical network resource capacity (e.g., when oversubscription is desirable).

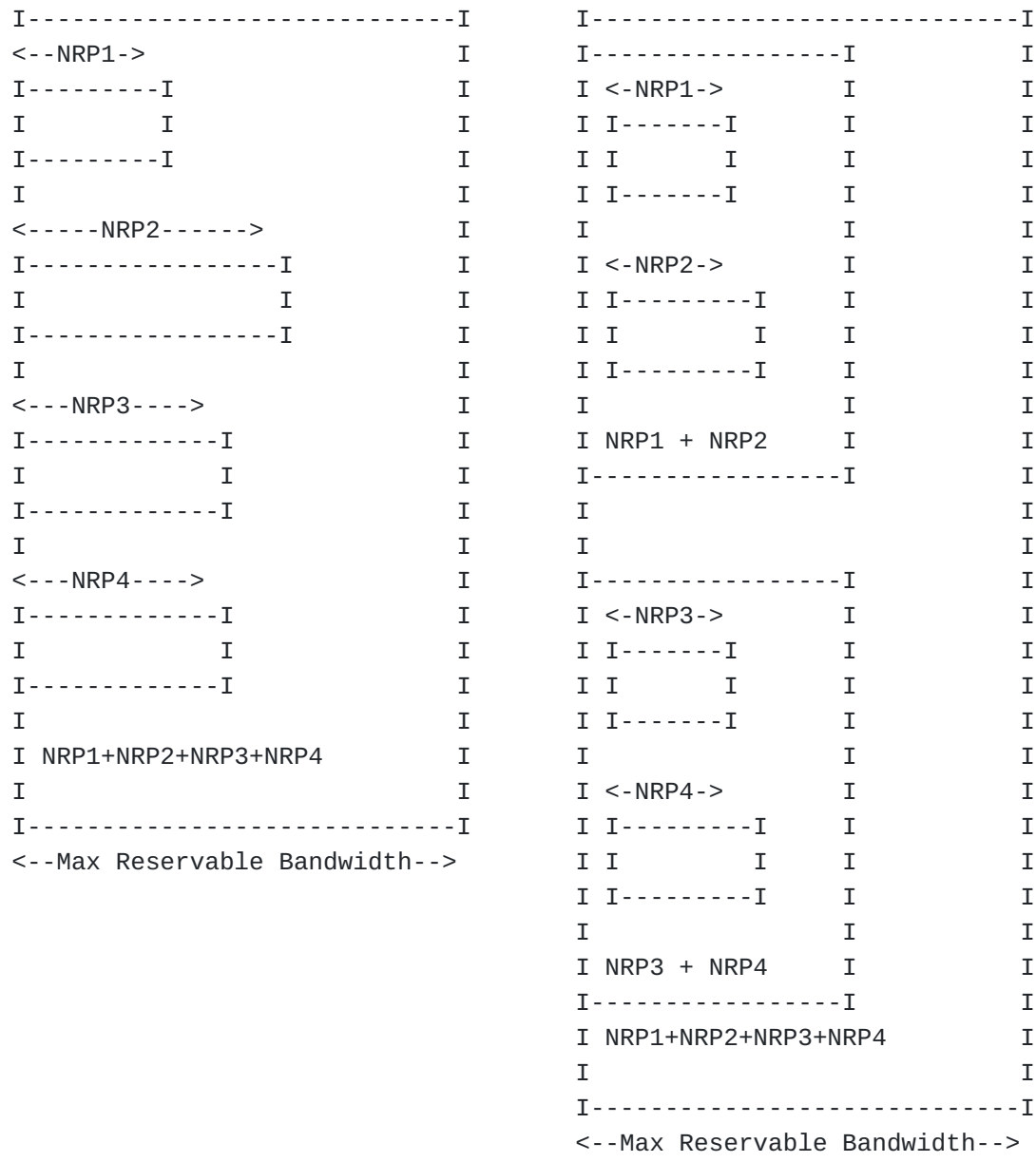
For example, the physical link bandwidth can be divided into fractions, each dedicated to an NRP that supports a Slice-Flow Aggregate. The topology associated with the NRP supporting a Slice-Flow Aggregate can be used by routing protocols, or by the ingress/ PCE when computing NRP state aware TE paths.

To perform NRP state aware Traffic Engineering (NRP-TE), the resource reservation on each link needs to be NRP aware. The NRP reservations state can be managed locally on the device or off device (e.g. on a controller).

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

Multiple NRPs that support Slice-Flow Aggregates can form a group and share the available network resources allocated to each. In this case, a node can update the reservable bandwidth for each NRP to take into consideration the available bandwidth from other NRPs in the same group.

For illustration purposes, Figure 2 describes bandwidth partitioning or sharing amongst a group of NRPs. In Figure 2a, the NRPs identified by the following NRP-IDs: NRP1, NRP2, NRP3 and NRP4 are not sharing any bandwidths between each other. In Figure 2b, the NRPs: NRP1 and NRP2 can share the available bandwidth portion allocated to each amongst them. Similarly, NRP3 and NRP4 can share amongst themselves any available bandwidth allocated to them, but they cannot share available bandwidth allocated to NRP1 or NRP2. 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 NRPs.

(b) Sharing bandwidth between NRPs of the same group.

Figure 2: Bandwidth isolation/sharing among NRPs.

4.3. Data and Control Plane Network Resource Partition Mode

In order to support strict guarantees for Slice-Flow 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 NRP that each supports a Slice-Flow Aggregate. The ingress routers or a Path Computation Engine (PCE) may use the

customized topologies and the NRP state to determine optimal path placement for specific demand flows using NRP-TE.

The data plane partitioning provides isolation for Slice-Flow Aggregate traffic, and protection when resource contention occurs due to bursts of traffic from other Slice-Flow Aggregate traffic that traverses the same shared network resource.

5. Network Resource Partition Instantiation

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

The customer of a network slice service expresses their intent by specifying requirements rather than mechanisms to realize the slice as described in Section 3.2.

The network slice controller is fed with the network slice service intent and realizes it with an appropriate Network Resource Partition Policy (NRP Policy). Multiple IETF network slices are mapped to the same Slice-Flow Aggregate as described in Section 3.3.

The network wide consistent NRP Policy definition is distributed to the devices in the network as shown in Figure 1. 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-Flow Aggregate are outside the scope of this document and will be addressed in separate documents.

5.1. NRP Policy Definition

The NRP Policy is network-wide construct that is supplied to network devices, and may include rules that control the following:

* Data plane specific policies: This includes the FAS, any firewall rules or flow-spec filters, and QoS profiles associated with the NRP Policy and any classes within it.

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

* Topology membership policies: This defines the topology filter policies that dictate node/link/function membership to a specific NRP.

There is a desire for flexibility in realizing network slices to support the services across networks consisting of implementations 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 NRP Policy will facilitate the instantiation and management of the NRP on the topological elements selected by the NRP Policy topology filter.

It is also possible to distribute the NRP 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 an NRP Policy definition will be described in separate documents.

5.1.1. Network Resource Partition - Flow-Aggregate Selector

A router should be able to identify a packet belonging to a Slice-Flow Aggregate before it can apply the associated dataplane forwarding treatment or NRP-PHB. One or more fields within the packet are used as an FAS to do this.

Forwarding Address Based FAS:

It is possible to assign a different forwarding address (or MPLS forwarding label in case of MPLS network) for each Slice-Flow 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-Flow Aggregate allows Slice-Flow 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-Flow 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-Flow 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 FAS:

An NRP Policy may include a Global Identifier FAS (G-FAS) field that is carried in each packet in order to associate it to the NRP supporting a Slice-Flow Aggregate, independent of the forwarding address or MPLS forwarding label that is bound to the destination. Routers within the NRP domain can use the forwarding address (or MPLS forwarding label) to determine the forwarding next-hop(s), and use the G-FAS field in the packet to infer the specific forwarding treatment that needs to be applied on the packet.

The G-FAS can be carried in one of multiple fields within the packet, depending on the dataplane used. For example, in MPLS networks, the G-FAS 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-Flow Aggregate may carry the same G-FAS in the MPLS label stack. It is also possible to have multiple G-FAS's map to the same Slice-Flow Aggregate.

The G-FAS 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 G-FAS to allow VPN packets to be mapped to the Slice-Flow Aggregate. In this case, a single VPN service label acting as a G-FAS may be allocated by all Egress PEs of a VPN. Alternatively, multiple VPN service labels may act as G-FAS's that map a single VPN to the same Slice-Flow 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 G-FAS's may map multiple VPN traffic to a single Slice-Flow Aggregate. An example of such deployment is shown in Figure 3.

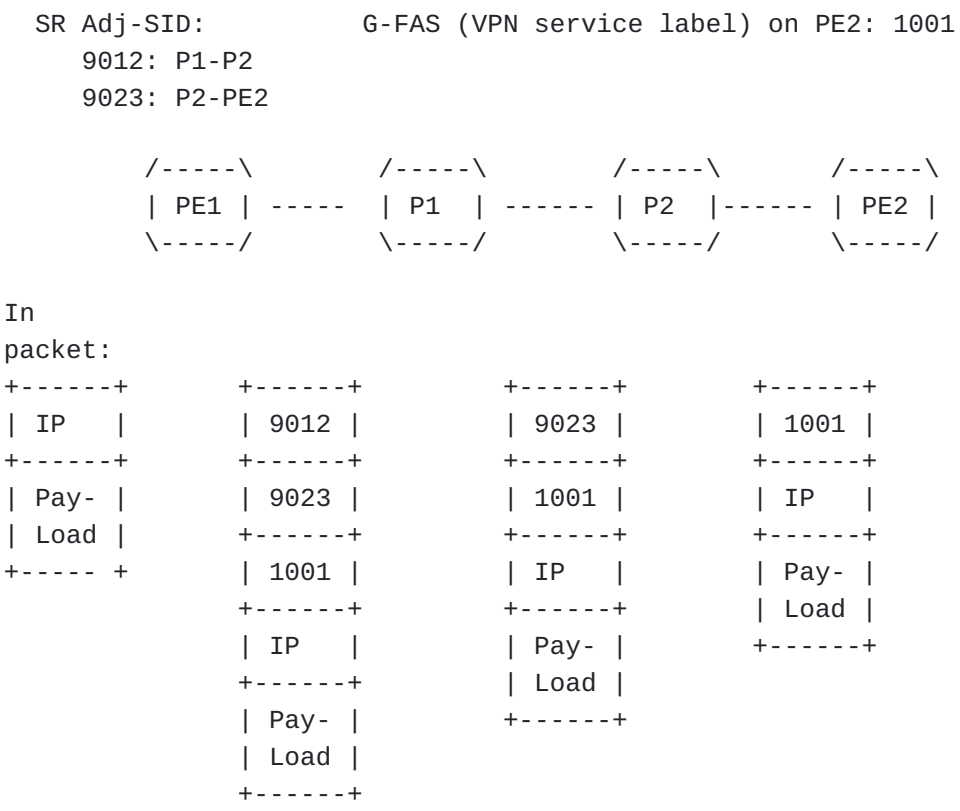


Figure 3: G-FAS or VPN label at bottom of label stack.

In some cases, the position of the G-FAS may not be at a fixed position in the MPLS label header. In this case, the G-FAS label can show up in any position in the MPLS label stack. To enable a transit router to identify the position of the G-FAS label, a special purpose label can be used to indicate the presence of a G-FAS in the MPLS label stack as shown in Figure 4.

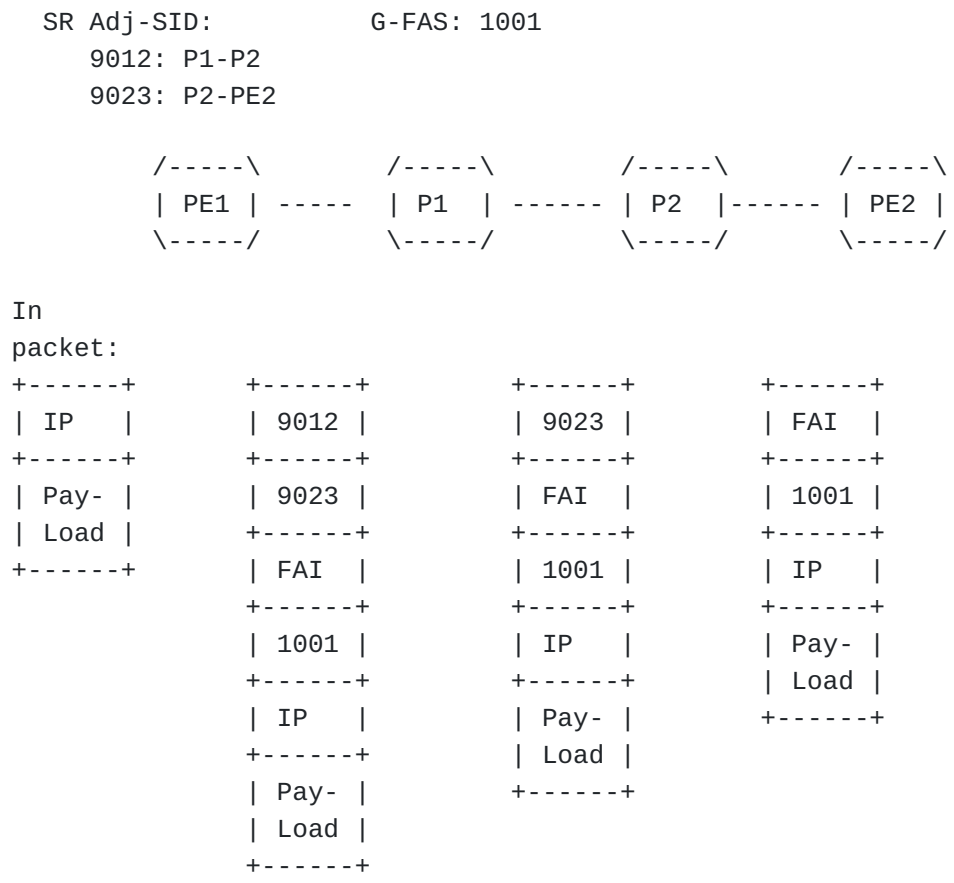


Figure 4: FAI and G-FAS label in the label stack.

When the slice is realized over an IP dataplane, the G-FAS can be encoded in the IP header (e.g. as an IPv6 option header).

5.1.2. Network Resource Partition 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 management of available bandwidth and the 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 NRPs may also be desirable. For example, a Slice-Flow Aggregate may not be using all of the NRP reservable bandwidth; this allows other NRPs in the same group to use the available bandwidth resources for other Slice-Flow Aggregates.

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

Since network characteristics and its state can change over time, the NRP topology and its network state 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 NRP network resources.

5.1.3. Network Resource Partition 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.

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

The Slice-Flow Aggregate traffic may be identified at NRP ingress boundary nodes by carrying a FAS 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 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-Flow Aggregate, a Slice-Flow 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 NRP.

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. Network Resource Partition Topology

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

An NRP topology can overlap or share a subset of links with another NRP topology. A number of topology filtering policies can be defined as part of the NRP Policy to limit the specific topology elements

that belong to the NRP. For example, a topology filtering policy can leverage Resource Affinities as defined in [RFC2702] to include or exclude certain links that the NRP is instantiated on in supports of the Slice-Flow Aggregate.

The NRP 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. Network Resource Partition Boundary

A network slice originates at the edge nodes of a network slice provider. Traffic that is steered over the corresponding NRP supporting a Slice-Flow Aggregate may traverse NRP capable as well as NRP 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. Network Resource Partition Edge Nodes

NRP edge nodes sit at the boundary of a network slice provider network and receive traffic that requires steering over network resources specific to a NRP that supports a Slice-Flow Aggregate. These edge nodes are responsible for identifying Slice-Flow 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 NRP 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 NRP mode is employed, the NRP ingress nodes are responsible for adding a suitable FAS onto packets that belong to specific Slice-Flow Aggregate. In addition, edge nodes may mark the corresponding Diffserv CS to differentiate between different types of traffic carried over the same Slice-Flow Aggregate.

5.2.2. Network Resource Partition Interior Nodes

An NRP interior node receives slice traffic and may be able to identify the packets belonging to a specific Slice-Flow Aggregate by

inspecting the FAS 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-Flow Aggregate. For example, when data plane NRP mode is applied, interior nodes can use the FAS carried within the packet to apply the corresponding NRP-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 NRP Policy, and allow differentiation of forwarding treatments for packets forwarded over the same NRP that supports the Slice-Flow Aggregate.

5.2.3. Network Resource Partition Incapable Nodes

Packets that belong to a Slice-Flow Aggregate may need to traverse nodes that are NRP 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 NRP forwarding treatment once the traffic reaches devices that are NRP-capable.

When data plane NRP mode is employed, packets carry a FAS to allow slice interior nodes to identify them. To support end-to-end network slicing, the FAS is maintained in the packets as they traverse devices within the network - including NRP capable and incapable devices.

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

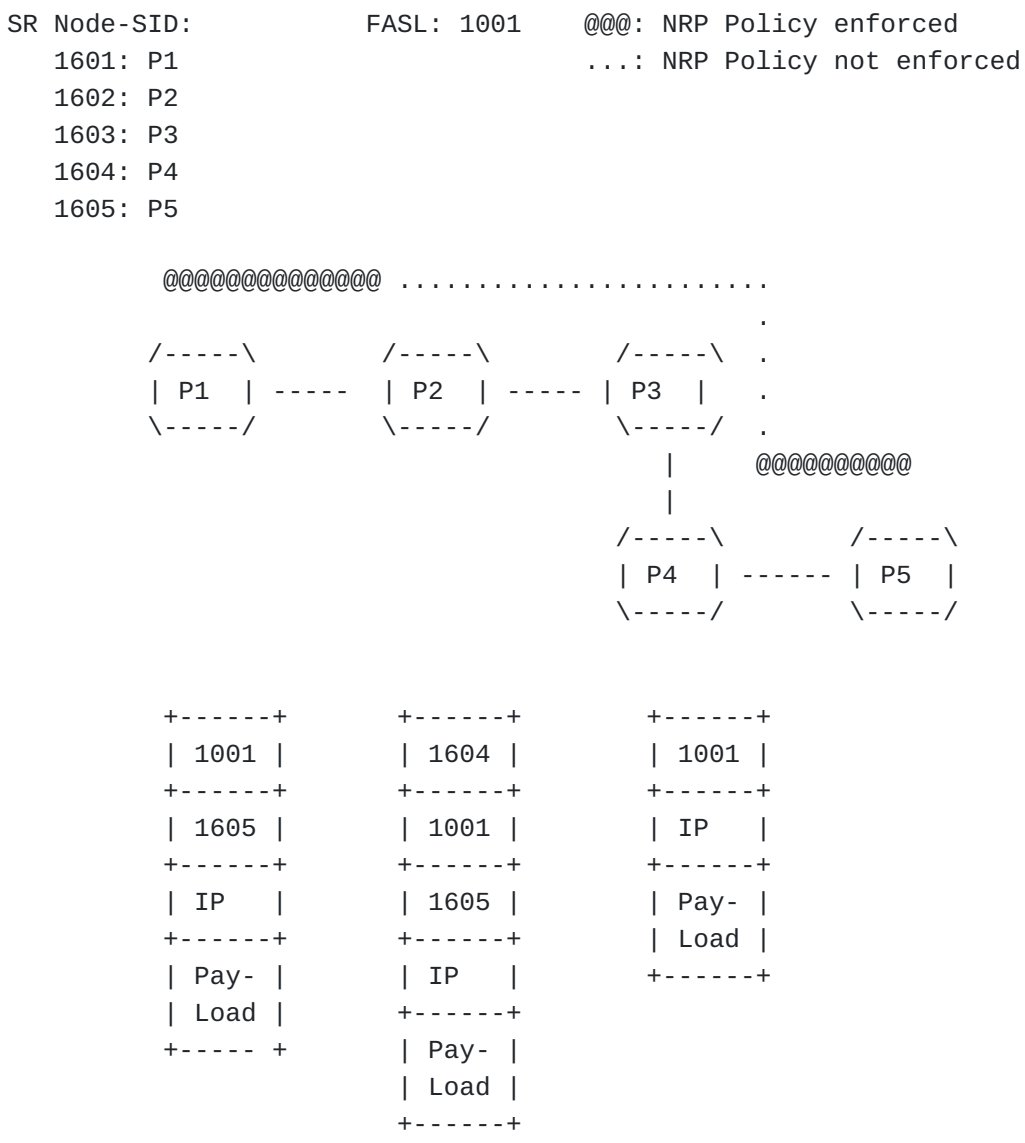


Figure 5: Extending network slice over NRP incapable device(s).

5.2.4. Combining Network Resource Partition Modes

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

6. Mapping Traffic on Slice-Flow Aggregates

The usual techniques to steer traffic onto paths can be applicable when steering traffic over paths established for a specific Slice-Flow Aggregate.

For example, one or more (layer-2 or layer-3) VPN services can be directly mapped to paths established for a Slice-Flow 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-Flow Aggregate path. External interfaces can be further partitioned (e.g., using VLANs) to allow mapping one or more VLANs to specific Slice-Flow 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 belongs to a specific Slice-Flow Aggregate. This option allows for applying a rich set of rules to identify specific packets to be mapped to a Slice-Flow Aggregate. However, it requires data plane network resources to be able to perform the additional checks in hardware.

6.1. Network Slice-Flow Aggregate Relationships

The following describes the generalization relationships between the IETF network slice and different parts of the solution as described in Figure 1.

o A customer may request one or more IETF Network Slices.

o Any given Attachment Circuit (AC) may support the traffic for one or more IETF Network Slices. If there is more than one IETF Network Slice using a single AC, the IETF Network Slice Service request must include enough information to allow the edge nodes to demultiplex the traffic for the different IETF Network Slices.

o By definition, multiple IETF Network Slices may be mapped to a single Slice-Flow Aggregate. However, it is possible for an Slice-Flow Aggregate to contain just a single IETF Network Slice.

o The physical network may be filtered to multiple Filter Topologies. Each such Filter Topology facilitates planning the placement of paths for the Slice-Flow Aggregate by presenting only the subset of links and nodes that meet specific criteria. Note,

however, in absence of any Filter Topology, Slice-Flow Aggregate are free to operate over the full physical network.

o It is anticipated that there may be very many IETF Network Slices supported by a network operator over a single physical network. A network may support a limited number of Slice-Flow Aggregates, with each of the Slice-Flow Aggregates grouping any number of the IETF Network Slices streams.

7. Path Selection and Instantiation

7.1. Applicability of Path Selection to Slice-Flow Aggregates

In State-dependent TE [I-D.ietf-teas-rfc3272bis], the path selection adapts based on the current state of the network. The state of the network can be based on parameters flooded by the routers as described in [RFC2702]. 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). TE extensions exist today to allow IGPs (e.g., [RFC3630] and [RFC5305]), and BGP-LS [RFC7752] to advertise such link state reservations.

When the network resource reservations are maintained for NRPs, the link state can carry per NRP state (e.g., reservable bandwidth). This allows path computation to take into account the specific network resources available for an NRP. In this case, we refer to the process of path placement and path provisioning as NRP aware TE (NRP-TE).

7.2. Applicability of Path Control Technologies to Slice-Flow Aggregates

The NRP modes described in this document are agnostic to the technology used to setup paths that carry Slice-Flow 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 NRP that supports a Slice-Flow Aggregate.

The feasible paths can be computed using the NRP topology and network state subject the optimization metrics and constraints.

7.2.1. RSVP-TE Based Slice-Flow Aggregate Paths

RSVP-TE [RFC3209] can be used to signal LSPs over the computed feasible paths in order to carry the Slice-Flow Aggregate traffic. The specific extensions to the RSVP-TE protocol required to enable

signaling of NRP aware RSVP-TE LSPs are outside the scope of this document.

7.2.2. SR Based Slice-Flow Aggregate Paths

Segment Routing (SR) [RFC8402] can be used to setup and steer traffic over the computed Slice-Flow Aggregate feasible paths.

The SR architecture defines a number of building blocks that can be leveraged to support the realization of NRPs that support Slice-Flow Aggregates in an SR network.

Such building blocks include:

- *SR Policy with or without Flexible Algorithm.
- *Steering of services (e.g. VPN) traffic over SR paths
- *SR Operation, Administration and Management (OAM) and Performance Management (PM)

SR allows a headend node to steer packets onto specific SR paths using a Segment Routing Policy (SR Policy). The SR policy supports various optimization objectives and constraints and can be used to steer Slice-Flow Aggregate traffic in the SR network.

The SR policy can be instantiated with or without the IGP Flexible Algorithm (Flex-Algorithm) feature. It may be possible to dedicate a single SR Flex-Algorithm to compute and instantiate SR paths for one Slice-Flow Aggregate traffic. In this case, the SR Flex-Algorithm computed paths and Flex-Algorithm SR SIDs are not shared by other Slice-Flow Aggregates traffic. However, to allow for better scale, it may be desirable for multiple Slice-Flow Aggregates traffic to share the same SR Flex-Algorithm computed paths and SIDs.

8. Network Resource Partition Protocol Extensions

Routing protocols may need to be extended to carry additional per NRP 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 an NRP Policy may need to be automated. Multiple options are possible to facilitate automation of distribution of an NRP Policy to capable devices.

For example, a YANG data model for the NRP 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 NRP Policy YANG data model is outside the scope of this document.

9. Outstanding Issues

Note to RFC Editor: Please remove this section prior to publication.

This section records non-blocking issues that were raised during the Working Group Adoption Poll for the document. The below list of issues needs to be fully addressed before progressing the document to publication in IESG.

1. Add new Appendix section with examples for the NRP modes described in Section 4.

2. Add text to clarify the relationship between Slice-Flow Aggregates, the NRP Policy, and the NRP.

3. Remove redundant references to Diffserv behaviors.

4. Elaborate on the SFA packet treatment when no rules to associate the packet to an NRP are defined in the NRP Policy.

5. Clarify the NRP instantiation through the NRP Policy enforcement.

6. Clarify how the solution caters to the different IETF Network Slice Service Demarcation Point locations described in Section 4.2 of [I-D.ietf-teas-ietf-network-slices].

7. Clarify the relationship the underlay physical network, the filter topology and the NRP resources.

8. Expand on how isolation between NRPs can be realized depending on the deployed NRP mode.

9. Revise Section 5.2.3 to describe how nodes can discover NRP incapable downstream neighbors.

10. Expand Section 11 on additional security threats introduced with the solution.

11. Expand Section 5.2 on NRP domain boundary and multi-domain aspects.

10. IANA Considerations

This document has no IANA actions.

11. 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 NRP is indicated primarily by the FAS, and hence an adversary may be able to utilize resources allocated to a specific NRP by injecting packets carrying the same FAS 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 NRP.

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

12. Acknowledgement

The authors would like to thank Krzysztof Szarkowicz, Swamy SRK, Navaneetha Krishnan, Prabhu Raj Villadathu Karunakaran, and Mohamed Boucadair for their review of this document and for providing valuable feedback on it. The authors would also like to thank Adrian Farrel for detailed discussions that resulted in Section 3.

13. 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

Adrian Farrel
Old Dog Consulting
United Kingdom
Email: adrian@olddog.co.uk

14. References

14.1. Normative References

[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>.

[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>.

[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>.

14.2. Informative References

[I-D.ietf-lsr-flex-algo] Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", Work in Progress, Internet-Draft, draft-ietf-lsr-flex-

algo-20, 18 May 2022, <https://www.ietf.org/archive/id/draft-ietf-lsr-flex-algo-20.txt>.

[I-D.ietf-teas-ietf-network-slices]
Farrel, A., Drake, J., Rokui, R., Homma, S., Makhijani, K., Contreras, L. M., and J. Tantsura, "Framework for IETF Network Slices", Work in Progress, Internet-Draft, draft-ietf-teas-ietf-network-slices-10, 27 March 2022, <https://www.ietf.org/archive/id/draft-ietf-teas-ietf-network-slices-10.txt>.

[I-D.ietf-teas-rfc3272bis]
Farrel, A., "Overview and Principles of Internet Traffic Engineering", Work in Progress, Internet-Draft, draft-ietf-teas-rfc3272bis-16, 24 March 2022, <https://www.ietf.org/archive/id/draft-ietf-teas-rfc3272bis-16.txt>.

[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>.

[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>.

[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>.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, DOI 10.17487/RFC5462, February 2009, <https://www.rfc-editor.org/info/rfc5462>.

[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>.

[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>.

Authors' Addresses

Tarek Saad
Juniper Networks

Email: tsaad@juniper.net

Vishnu Pavan Beeram
Juniper Networks

Email: vbeeram@juniper.net

Jie Dong
Huawei Technologies

Email: jie.dong@huawei.com

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
Ciena

Email: rrokui@ciena.com

Luay Jalil
Verizon

Email: luay.jalil@verizon.com