

RTGWG
Internet-Draft
Intended status: Standards Track
Expires: November 10, 2022

S. Bryant, Ed.
University of Surrey 5GIC
U. Chunduri, Ed.
Intel Corporation
A. Clemm
Futurewei
May 09, 2022

Preferred Path Routing Framework
draft-chunduri-rtgwg-preferred-path-routing-02

Abstract

Capacity demands, Traffic Engineering (TE) and determinism are some of key requirements for various cellular, edge and industrial deployments. These deployments span from many underlying data plane technologies including native IPv4, native IPv6 along with MPLS and Segment Routing (SR).

This document provides a framework for Preferred Path Routing (PPR). PPR is a method of providing path based dynamic routing for a number of packet types including IPv4, IPv6 and MPLS. This seamlessly works with a controller plane which holds both complete network view of operator policies, while distributed control plane providing self-healing benefits in a near-real time fashion.

PPR builds on existing encapsulations at the edge nodes to add path identity to the packet. This reduces the per packet overhead required for path steering and therefore has a smaller impact on both packet MTU, data plane processing and overall goodput for small payload packets, while extending path steering benefits to any existing data plane.

A number of extensions that allow expansion of use beyond simple point-to-point-paths are also described in this memo.

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 November 10, 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Relation to Segment Routing	4
1.2. Requirements Language	4
1.3. Acronyms	4
2. Applicability and Key use cases	5
2.1. XHaul Transport	6
2.2. PPR as VPN+ Underlay and Network Slicing	7
2.3. PPR as FRR Solution	7
2.4. Extensible alternative to Flex Algo	8
2.5. PPR for energy-optimized networks	8
3. Preferred Path Routing (PPR)	9
3.1. PPR Data Plane aspects	11
3.1.1. PPR Native IP Data Planes	11
3.1.2. SR-MPLS with PPR	13
3.1.3. SRv6, Network Programming and PPR	13
3.2. PPR Control Plane aspects	14
3.2.1. PPR-ID and Data Plane Extensibility	14
3.2.2. PPR Path Description Elements (PDEs)	14
3.2.3. ECMP Considerations	15
3.2.4. PPR Services along the Path	15
3.2.5. PPR Graphs	15
3.2.6. PPR Multi-Domain Scenarios	17
3.3. PPR Management Plane Aspects	18
3.3.1. IGP Metric Independent Paths/Graphs	18
3.3.2. Granular OAM	19
4. Preferred Path Loop Free Alternatives (pLFA)	20

5. Traffic Engineering Attributes	21
6. IANA Considerations	22
7. Security Considerations	22
8. References	22
8.1. Normative References	22
8.2. Informative References	22
Authors' Addresses	24

1. Introduction

With the deployments of more advanced services, such as 5G and beyond, the need for Traffic Engineering (TE) with deterministic services become more important. Especially in many edge networks where stringent requirements must be met in terms of latency, throughput, packet loss and packet error rate. Traffic steering provides a base to build some of these capabilities to serve various cellular, edge and vertical industries. Additionally, diverse data planes are used in various deployments and parts of the network, including Ethernet, MPLS, and native IP (IPv4/IPv6) needs some of these capabilities.

This document provides a framework for Preferred Path Routing (PPR). PPR is a method of adding explicit paths to a network using link-state routing protocols. Such a path, which may be a strict or loose and can be any loop-free path between two points in the network. A node makes an on-path check to determine if it is on the path, and, if so, adds a FIB entry with NextHop (NH) (computed from the SPF tree) set to the next element in the path description.

The Preferred Path Route Identifier (PPR-ID) in the packet is used to map the packet to the PPR path, and hence to identify resources and the NH. In other words, PPR-ID is the path identity to the packet and routing and forwarding happens based on this identifier while providing various services to all the flows mapped to the path.

As described, PPR is forwarding plane agnostic, and may be used with any packet technology in which the packet carries an identifier that is unique within the PPR domain. PPR may hence be used to add explicit path and resource mapping functionality with inherent TE properties in IPv4, IPv6, MPLS, Ethernet or similar networks. It also has a smaller impact on both packet MTU and data plane processing. PPR uses an IGP control plane based approach for dynamic path steering.

Applications and key use case scenarios for PPR are described further in Section 2.

PPR itself is described in further detail in Section 3.

Section 3.1, Section 3.2, Section 3.3 describe various data, control and management plane functionalities of PPR respectively. A number of extensions that allow expansion of use beyond simple point-to-point- paths, TE aware loop-free-alternatives and path related services to the flows are also described in this memo.

1.1. Relation to Segment Routing

Segment Routing (SR) [RFC8402] enables packet steering by including set of Segment Identifiers (SIDs) in the packet that the packet must traverse or be processed by. In an MPLS network this is done by mapping the SIDs to MPLS labels and then pushing the required labels on the packet [RFC8660]. In SRv6, [RFC8754] defines a segment routing extension header (SRH) to be carried in the packet which contains a list of the segments. The usefulness of PPR with SR and inter- working scenarios are described in Section 3.1.2 and Section 3.1.3.

SR also defines Binding SIDs (BSIDs) [RFC8402] which are SIDs pre-positioned in the network to either allow the number of SIDs in the packet to be reduced, or provide a method of translating from an edge imposed SID to a SID that the network prefers. One use of BSIDs is to define a path by associating an out-bound SID on every node along the path in which case the packet can be steered by swapping the incoming active SID on the packet with a BSID. For both SR-MPLS and SRv6, PPR can reduce the number of touch points needed with BSIDs by dynamically signaling the path and associating the path with an abstract data plane identifier.

With PPR as a data packet carries a PPR-ID Section 3.1 instead of individual SIDs, it avoids exposing the path; thus it avoids revealing topology, traffic flow and service usage, if a packet is snooped. This is described as "Topology Disclosure" security consideration in [RFC8754].

1.2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119 [RFC2119], RFC8174 [RFC8174] when, and only when they appear in all capitals, as shown here.

1.3. Acronyms

Term	Definition
GTP	GPRS Tunneling Protocol (3GPP)
IS-IS LSP	IS-IS Link State PDU
IPFRR	IP FastReRoute
MPLS	Multi-Protocol Label Switching
MTU	Maximum Transferable Unit
NH	NextHop
PDE	Path Description Element of the Preferred path
PE	Provider Edge
PPR	Preferred Path Routing/Route
PPR-ID	Preferred Path Route Identifier, a data plane identifier
(R)AN	5G Radio Access Network (3GPP REL15)
SID	Segment Identifier
SPF	Shortest Path First
SR-MPLS	Segment Routing with MPLS data plane
SRH	Segment Routing Header - IPv6 routing Extension header
SRv6	Segment Routing with IPv6 data plane with SRH
TE	Traffic Engineering
UPF	User Plane Function (3GPP REL15)

2. Applicability and Key use cases

2.1. XHaul Transport

Cellular networks predominantly use both IP and MPLS data planes in the transport part of the network. For the cellular transport to evolve for 5G, certain underlay network requirements need to be met (for slices other than enhanced Mobile Broadband (eMBB)). PPR is a mechanism to achieve this, as it provides dynamic path based routing and traffic steering for any underlying data plane (IPv4/IPv6/MPLS) used, without any additional control plane protocol in the network. PPR acts as an underlay mechanism in cellular XHaul (N3/N9 interfaces below) and hence can work with any overlay mechanism including GPRS Tunneling Protocol (GTP).

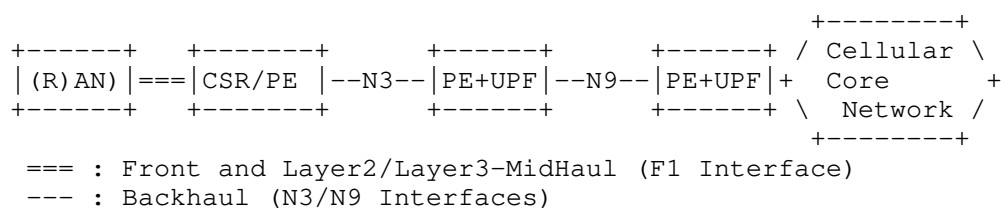


Figure 1: Cellular Transport Network

Figure 1 depicts a high level view of cellular XHaul network. Fronthaul is generally a point-to-point link, midhaul uses Layer-2/IP and backhaul is an IP/MPLS network. For the end-to-end slicing in these deployments, both midhaul and backhaul need to have traffic engineering as well as underlay QoS capabilities.

In many cellular deployments connectivity for various 5G nodes on F1, N3 and N9 interfaces, topologies for example, range from subtended rings to Leaf- Spine Fabric (LS-Fabric) in edge deployments. While there is no limitation w.r.t topologies where PPR is applicable, for some of these deployments PPR is more suitable for providing Traffic Engineering for high volume traffic with no path overhead in the underlying data plane. PPR augments the SR-MPLS deployments with low data plane overhead and high reliability with TE aware fast reroute (PLFA) as described in Section 3.2.2. In the overlay or virtual router environment, PPR provides lightweight service chaining with non-topological Path Description Element (PDEs) Section 3.2.2 along the preferred path. PPR helps to achieve OAM capabilities at the path granularity without any additional per packet information.

Some edge deployment underlays are predominantly IP (IPv4/IPv6) based. If IGP based underlay control plane is in use, PPR can provide the required flexibility for creating TE paths, where native IP data planes (IPv4/IPv6) are used. PPR can help operators to

mitigate the congestion in the underlay and path related services for critical servers in the edge networks dynamically.

2.2. PPR as VPN+ Underlay and Network Slicing

There is a need to support the requirements of new applications, particularly applications that are associated with 5G services. An approach to supporting these needs is described in [I-D.ietf-teas-enhanced-vpn]. This approach utilizes existing VPN and TE technologies and adds features that specific services require over and above traditional VPNs. The document describes a framework for using existing, modified and potential new networking technologies as components to provide an Enhanced Virtual Private Network (VPN+) service.

Typically, VPN+ will be used to form the underpinning of network slicing, but could also be of use in its own right. It is not envisaged that large numbers of VPN+ instances will be deployed in a network and, in particular, it is not intended that all VPNs supported by a network will use VPN+ techniques.

Such networks potentially need large numbers of paths each with individually allocated resources at each link and node. A segment routing approach has the potential to require large numbers of SIDs in each packet; and the paths become strict source routed through end-to-end set of resources needed to create the VPN+ paths. By using PPR, the number of segments needed in packets is reduced, and the management overhead of installing the large numbers of BSIDs is reduced.

2.3. PPR as FRR Solution

PPR may be used in a network as a method of providing IP Fast-ReRoute (IPFRR). This is independent of whether PPR is used in the network for other traffic steering purposes. It can be used to create optimal paths or paths congruent with the post convergence path from the point of local repair (PLR) as is proposed in TI-LFA [I-D.ietf-rtgwg-segment-routing-ti-lfa]. Unlike TI-LFA PPR may be used in IPv4 networks. This is discussed further in Section 4. The approach has the further intrinsic advantage that no matter how complex the repair path, only a single header (or MPLS label) needs to be pushed onto the packet which may assist routers that find it difficult to push large headers.

2.4. Extensible alternative to Flex Algo

Flex-Algorithm [I-D.ietf-lsr-flex-algo] is a method that is sometimes used to create paths between Segment Routing (SR) nodes when it is required that packets traverse a path other than the shortest path that the SPF of the underlying IGP would naturally install. There is a limit of 128 algorithms that can be installed in a network. Flex-Algorithm is a cost based approach to creating a path which means that a path or pathlet is indirectly created by manipulating the metrics of the links. These metrics affect all the paths within the scope of the Flex-Algorithm number (instance). The traffic steering properties of Flex-Algorithm required for SR can be achieved directly with PPR with several advantages:

- o The scope of a PPR path is strictly limited to the sub-path between the SR nodes.
- o The path can be directly specified rather than implicitly through metrics
- o Resources (such as specialist queues etc.) may be directly mapped to the PPR path and hence to the SR sub-path.

2.5. PPR for energy-optimized networks

Improving energy efficiency and reducing power consumption are becoming of increasing importance for many industries, which includes network operators as well as users and providers of networking services. Network providers can contribute to addressing those challenges by making their networks more energy-efficient. This includes the support of power saving schemes that guide traffic along paths deemed particularly "energy efficient".

A significant opportunity to reduce power consumption concerns powering down (or putting into power saving mode) equipment (including line card, ports, etc.) that may not be currently needed. At the same time, the incremental power consumption for additional traffic on ports and equipment already under power is for all practical purposes negligible. Optimizing energy efficiency thus involves directing traffic in such a way that it allows for isolation of equipment that may at the moment not be needed so that it could be powered down or brought into power-saving mode.

This implies that power efficiency of networks can be greatly affected by the paths along which traffic is directed at any particular point in time. Proper engineering of those paths and the ability to configure them effectively thus becomes an important tool to optimize power usage of networks and make them more energy

efficient. PPR provides a mechanism that enables this, allowing to engineer dynamic paths that optimize the network from an energy savings perspective as one important set of criteria for any underlying data plane in the network.

3. Preferred Path Routing (PPR)

PPR allows the direction of traffic along an engineered path through the network by replacing the SID label stack or the SID list with a single PPR-ID. The PPR-ID may either be a single label (MPLS) or a native destination prefix (IPv4/IPv6). This enables the use of a single data plane identifier to describe an entire path.

A PPR path could be an (Segmented Routed) SR path, a traffic engineered path computed based on some constraints, an explicitly provisioned Fast Re-Route (FRR) path or a service chained path. A PPR path can be signaled by any node, computed by a central controller, or manually configured by an operator. PPR extends the source routing and path steering capabilities to native IP (IPv4 and IPv6) data planes without hardware upgrades; see Section 3.1.

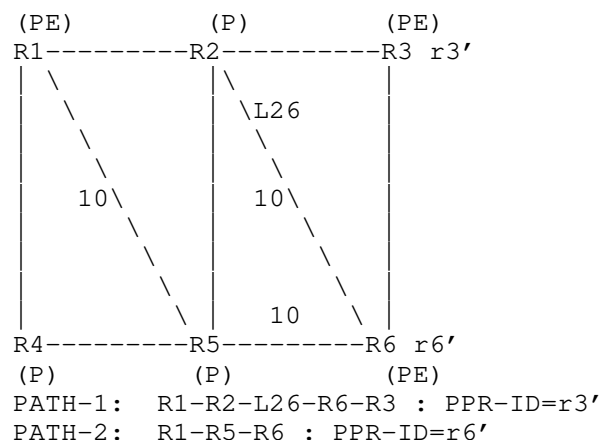


Figure 2: IGP Network

In the Figure 2, consider node R1 as an ingress node, or a head-end node, and the node R3 may be an egress node or another head-end node. The numbers shown on links between nodes indicate the bi-directional IGP metric as provisioned (no number indicates a metric of 1). R1 may be configured to receive TE source routed path information from a central entity (PCE [RFC5440], Netconf [RFC6241] or a Controller) that comprises PPR information which relates to sources that are attached to R1. It is also possible to have a PPR provisioned

locally by the operator for non-TE needs (e.g., FRR or for chaining certain services).

The PPR is encoded as an ordered list of path elements from source to a destination node in the network and is represented with a PPR-ID to represent the path. The path can represent both topological and non-topological elements (for example, links, nodes, queues, priority and processing actions) and specifies the actual path towards the egress node.

- o The shortest path towards R3 from R1 is through the following sequence of nodes: R1-R2-R3 based on the provisioned IGP metrics.
- o The central entity in this example, can define a PPRs from R1 to R3 and R1 to R6 that deviate from the shortest path based on other network characteristic requirements as requested by an application or service. For example, the network characteristics or performance requirements may include bandwidth, jitter, latency, throughput, error rate, etc.
- o In a VPN setup, nodes R1, R3 and R6 are PE nodes and other nodes are P nodes. User traffic entering at the ingress PE nodes gets encapsulated (e.g., MPLS, GRE, GTP, IP-IN-IP, GUE) and will be delivered to the egress PE.

Consider two paths in the above network:

- o PATH-1: A first PPR may be identified by PPR-ID = r3' with the path description R1-R2-L26-R6-R3 for a Prefix advertised by R3. This is an example of a strict path with a combination of links and nodes.
- o PATH-2: A second PPR may be identified by PPR-ID = r6' with the path description R1-R5-R6. This is an example of a loose path. Though this example shows PPRs with node identifiers it is possible to have a PPR with a combination of Non-Topological elements along the path.

The first topological element relative to the beginning of PPR Path descriptor contains the information about the first node in the path that the packet must pass through (e.g. equivalent to the top label in SR-MPLS and the first SID in an SRv6 SRH). The last topological sub-object or PDE contains information about the last node (e.g. in SR-MPLS it is equivalent to the bottom SR label).

Each IGP node receiving a complete path description, determines whether the node is on the advertised PPR path. This is called the PPR on-path check. It then determines whether it is included more

than once on that path. This PPR validation prevents the formation of a routing loop. If the path is looped, no further processing of the PPR is undertaken. (Note that even if it is invalid, the PPR descriptor must still be flooded to preserve the consistency of the underlying routing protocol). If the validation succeeds, the receiving IGP node installs a Forwarding Information dataBase (FIB) entry for the PPR-ID with the NextHop (NH) required to take the packet to the next topological path element in the path description. Processing of PPRs may be done at the end of the IGP SPF computation.

Consider PPR path PATH-1 in Figure 2. When node R5 receives the PPR (PATH-1) information it does not install a FIB entry for PATH-1 because this PPR does not include node R5 in the path description/ordered path list.

However, node R5 determines that the second PPR (PATH-2), does include the node R5 in its path description (the on-path check passes). Therefore, node R5 updates its FIB to include an entry for the destination address that R6 indicates (PPR-ID) along with path description. This allows the forwarding of data packets with the PPR-ID (r6') to the next element along the path, and hence towards node R6.

To summarize the control plane processing, the receiving IGP node determines if it is on the path by checking the node's topological elements in the path list. If it is, it adds/adjusts the PPR-ID's shortest path NH towards the next topological path element in the PPR's path list. This process continues at every IGP node as specified in the path description TLV.

3.1. PPR Data Plane aspects

Data plane type for PPR-ID is selected by the entity (e.g., a controller, locally provisioned by operator), which selects a particular PPR in the network.

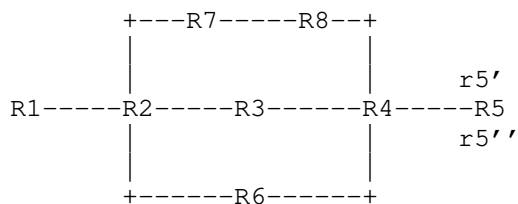
3.1.1. PPR Native IP Data Planes

In an IPv4 network, source routing and packet steering with PPR can be done by selecting the IPv4 data plane type (PPR-IPv4), in PPR Path description with a corresponding IPv4 address/prefix as PPR-ID while signaling the path description in the control plane Section 3.2. Forwarding is done by setting the destination IP address of the packet as PPR-ID at the ingress node of the network. In this case this is an IPv4 address in the tunneled/encapsulated user packet. There is no data plane change or upgrade needed to support this functionality.

Similarly, for an IPv6 network source routing and packet steering can be done in IPv6 data plane type (PPR-IPv6), along the path as described in PPR Path description with a corresponding IPv6 address/prefix as PPR-ID in the control plane Section 3.2. Whatever specified above for IPv4 applies here too, except that destination IP address of the encapsulated data packet at the edge node is an IPv6 Address (PPR-ID). This doesn't require any IPv6 extension headers (EH).

For a loose path in an IPv4 or IPv6 network (Native IPv4 or IPv6 data planes respectively) the packet has to be encapsulated using the capabilities (either dynamically signaled through [I-D.ietf-isis-encapsulation-cap] or statically provisioned on the nodes) of the next loose PDE in the path description.

Consider the network fragment shown in Figure 3 which further illustrates loose routing, and consider PATH-3. Node R2 can reach R5 ECMP through R2->R3->R4, and R2->R6->R4, both at cost 2. The path R2->R7->R8->R4 is longer (cost 3) and is not a path that R2 would choose to use it to reach R4. Node R2 (start of the loose segment) is programmed to encapsulate a data packet towards the next loose topological PPR-PDE in the path, which is R4. The NH computed at R1 (for PPR-ID r5') would be the shortest path towards R5 i.e., the interfaces towards R2. R2 has an ECMP towards R3 and R6 to reach R4 (next PDE in the loose segment), as packet would be encapsulated at R2 for R4 as the destination. R7 and R8 are not involved in this PPR path and so do not need a FIB entry for PPR-ID r5' (the on-path check for PATH-3 fails at these nodes).



All costs are 1

PATH-3: R1-R2-R4-R5 : PPR-ID=r5'

PATH-4: R1-R2-R3-R4-R5 : PPR-ID=r5''

Figure 3: Network with Loose Path

In a strict path, for example, PATH-4 in Figure 3, PPR-ID is programmed on the data plane at each node of the path, with NH set to the shortest path towards the next topological PPR-PDE. In this case, no further encapsulation of the data packet is required.

3.1.2. SR-MPLS with PPR

PPR is fully backward compatible with the SR data plane. As control plane PDEs can be extensible and particular data plane identifiers can be expressed to describe the path, in SR case PDEs can contain the SR SIDs.

In SR-MPLS, a data packet contains the stack of labels (path steering instructions) which guides the packet traversal in the network. For SR-MPLS data plane, the complete set of label stack is represented with a unique SR SID/Label, PPR-ID, to represent the path. The PPR-ID gets programmed on the data plane of each node, with the appropriate NH computed as specified in Section 3. PPR-ID here is a label/index from the SRGB (like another node SID or global ADJ-SID). PPR path description in the control plane is a set of ordered SIDs represented with PPR-PDEs. Non-Topological segments described along with the topological PDEs can also be programmed in the forwarding plane to enable specific function/service, when the data packet hits with corresponding PPR-ID.

For SR-MPLS data plane, either 1 label or 2 labels need to be provisioned on individual nodes on the path description. In the example network Figure 2, for PATH-2 (a loose path), during control plane processing, node R1 programs the bottom label as PPR-ID and the top label as the next topological PPR-PDE in the path, which is a node SID of R5. In the control plane, the NH computed at R1 would be the shortest path towards R5 i.e., the interfaces towards R2 and R4 (ECMP). For strict paths, a single label (PPR-ID) is programmed on the data plane along the path, with NH set to the shortest path towards the next topological PPR-PDE in the path description.

3.1.3. SRv6, Network Programming and PPR

One of the key benefits PPR offers for SRv6 data plane is an optimized data plane as individual path steering SIDs in the data packet is replaced with a path identifier (PPR-ID). Thus potentially avoids MTU, hardware incompatibilities and processing overhead. Few PPR and SRv6 inter working scenarios are listed below.

In a simple encapsulation mode without SRH [RFC8754], an SRv6 SID can be used as PPR-ID. With this approach path steering can be brought in with PPR and some of the network functions as defined in [RFC8986] can be realized at the egress node as PPR-ID in this case is a SRv6 SID.

In SRv6 with SRH, one-way PPR-ID can be used, by setting it as the destination IPv6 address and SL field in SRH is set to 0; here, SRH can contain any other TLVs and non-topological SIDs as needed.

Another inter working case can be a multi area IGP deployment. In this case multiple PPR-IDs corresponding to each IGP area can be encoded as SIDs in SRH for an end-to-end path steering with minimal SIDs in SRH.

3.2. PPR Control Plane aspects

3.2.1. PPR-ID and Data Plane Extensibility

The data plane identifier, PPR-ID, describes a path through the network. A data plane type and corresponding PPR-ID can be specified with the advertised path description in the IGP. The PPR-ID type allows data plane extensibility for PPR, though it is currently defined for IPv4, IPv6, SR-MPLS and SRv6 data planes.

For native IP data planes, this is mapped to either IPv4 or IPv6 address/prefix. For SR-MPLS, PPR-ID is mapped to an MPLS Label/SID and for SRv6, this is mapped to an IPv6-SID. This is further detailed in Section 3.1 and Section 3.1.3.

3.2.2. PPR Path Description Elements (PDEs)

The path identified by the PPR-ID is described as a set of Path Description Elements (PDEs), each of which represents a segment of the path. Each node determines its location in the path as described, and forwards to the next segment/hop or label of the path description (see the Forwarding Procedure Example later in this document).

These PPR-PDEs like SR SIDs, can represent topological elements like links/nodes, backup nodes, as well as non- topological elements such as a service, function, or context on a node with additional control information as needed.

A preferred path can be described as a Strict-PPR or a Loose-PPR. In a Strict-PPR all nodes/links on the path are described with SR-SIDs for SR data planes or IPv4/IPV6 addresses for native IP data planes. In a Loose-PPR only some of the nodes/links from source to destination are described. More specifics and restrictions around Strict/Loose PPRs are described in respective data planes in Section 3.1 and Section 3.1.3. Each PDE is described as either an MPLS label towards the NH in MPLS enabled networks, or as an IP NH, in the case of either 'plain'/'native' IP or SRv6 enabled networks. A PPR path is related to a set of PDEs using the TLVs specified in the respective IGPs.

3.2.3. ECMP Considerations

PPR inherently supports Equal Cost Multi Path (ECMP) for both strict and loose paths. If a path is described using nodes, it would have ECMP NHs established for PPR-ID along the path. In the network shown in Figure 2, for PATH-2, node R1 would establish ECMP NHs computed by the IGP, towards R5 for the PPR-ID r6'. However, one can avoid ECMP on any segment of the path by pinning the path using link identifier to the next segment as specified for PATH-1 in Figure 2.

3.2.4. PPR Services along the Path

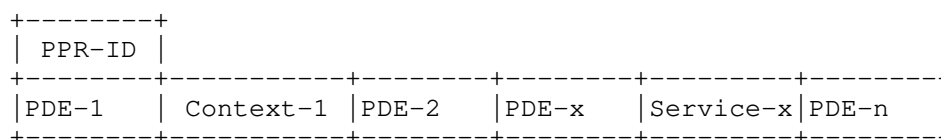


Figure 4: Services along the Preferred Path

As shown in Figure 4, some of the services specific to a preferred path, can be encoded as non-topological PDEs and can be part of the path description. These services are applied at the respective nodes along the path. In Figure 4, PDE-1, PDE-2, PDE-x, PDE-n are topological PDEs of a data plane. For SR-MPLS/SRv6 data planes these are simply SIDs and for native IP data planes corresponding non-topological addresses. When the data packet with a PPR-ID is delivered to node-1, the packet is delivered to Context-1. Similarly on node-x, Service-x is applied. These services/functions need to be pre-provisioned on the particular nodes and optionally can be advertised in IGPs.

The above gives the basic and light weight service chaining capability with PPR without incurring any additional overhead on the data packet. However, this is limited to fixed services/functions for a path and all data packets using the path will be applied with these services. Flow level exclusions using the same path or differentiated services that need to be applied with in a flow cannot be supported with this mechanism and one has to resort to data plane mechanisms as defined in NSH/SFC [RFC8300].

3.2.5. PPR Graphs

In a network of N nodes a total $O(N^2)$ unidirectional paths are necessary to establish any-to-any connectivity, and multiple (k) such path sets may be desirable if multiple path policies are to be supported (lowest latency, highest throughput etc.).

In many solutions and topologies, N may be small enough and/or only a small set of paths need to be preferred paths, for example for high value traffic (DetNet, some of the defined 5G slices), and then a point-to-point path structure specified in this document can support these deployments.

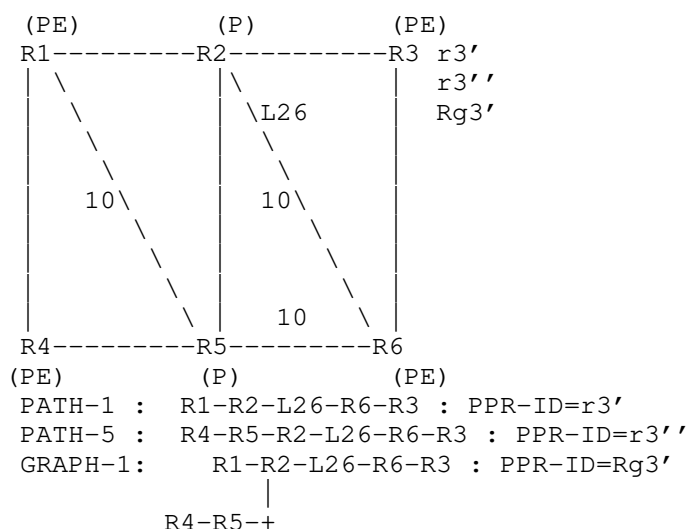


Figure 5: Network with a Graph Structure: PPR TREE

However, to address the scale needed when a larger number of PPR paths are required and for TE aware IPFRR Section 4, PPR TREE structure can be used.

Consider the network fragment in Figure 5, where two PPR paths, PATH-1 and PATH-5 are shown from different ingress PE nodes (R1, R4) to the same egress PE node (R3). In a simple PPR Tree structure, these 2 paths can be combined to form a PPR Tree structure. PPR Tree is one type of a graph where multiple source nodes are rooted at one particular destination node, with one or more branches. Figure 5, shows a PPR TREE (GRAPH-1), with 2 branches constructed with different PDEs, has a common PDE (node R2) and with a forwarding Identifier Rq3' (PPR-ID) at the destination node R3.

Each PPR Tree uses one label/SID and defines paths from any set of nodes to one destination, this reduces the number of entries needed. For example, it reduces the number of forwarding identifiers needed in SR-MPLS data plane Section 3.1.2 with PPR, which are derived from the SRGB at the egress node. These paths are form a tree rooted at the destination. In other word, PPR Tree identifiers are destination identifiers and PPR Trees are path engineered destination routes

(like IP routes) and its scaling simplifies to linear in N i.e., $O(k*N)$.

In a completely different usage paradigm, a PPR Graph can also have multiple forwarding identifiers (PPR-IDs). Based on the algorithm specified for the Graph, path computation can be done in a distributed fashion in the network to establish the forwarding over the graph. Various types of PPR Graphs, rules for construction and their usage details will be described in future revisions.

3.2.6. PPR Multi-Domain Scenarios

PPR can be extended to multi-domain, including multi-area scenarios as shown in Figure 6.

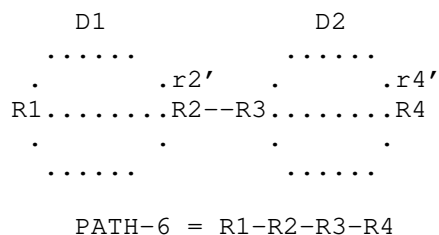


Figure 6: Multi-Domain Network with PPR

Operation of PPR within the domain is as described in the preceding sections of this document. The key difference in operation in multi-domain concerns the value of the PPR-ID in the packet. There are three approaches that can be taken:

1. The PPR-ID is constant along the end-end-path. This requires coordination of the PPR-ID in each domain. This has the convenience of a uniform identity for the path. However, whilst an IPv6 network has a large PPR identity space, this is not the case for MPLS and is less the case for IPv4. The approach also has the disadvantage that the entirety of the domains involved need to be configured and provisioned with the common value. In the network shown in Figure 6 The PPR-ID for PATH-6 is $r4'$.
2. The PPR-ID for each individual domain is the value that best suits that domain, and the PPR-ID is swapped at the boundary of the domains. This allows a PPR-ID that best suits each domain. This is similar to the approach taken with multi-segment pseudowire [RFC5659]. This approach better suits the needs of network layers with limited identity resources. It also enables the better coordination of PPR-IDs. In this approach the PPR-ID for PATH-6 would be $r2'$ in domain D1 and $r4'$ in domain D2. These

two PPR-IDs would be distributed in their own domains and the only inter-domain co-ordination required would be between R2 and R3.

3. A variant of (2) is that the PPR-IDs are domain specific, but a segment routing approach is taken in which they are encoded at ingress (R1), and are popped at the inter-domain boarder. This requires that the domain ingress and egress routers support segment routing data-plane capability.

Although the example shown in Figure 6 shows the case of two domains, nothing limits the design to just two IGP areas. This is further explained below.

In controller based deployments, each IGP area can have separate north bound and south bound communication end points with PCE/SDN controller, in their respective domain. It is expected that PPR paths for each IGP level are computed and provisioned at the ingress nodes of the corresponding area's area boarder router. Separate path advertisement in the respective IGP area should happen with the same PPR-ID. With this, only PPR-ID needs to be leaked to the other area, as long as a path is available in the destination area for that PPR-ID. If the destination area is not provisioned with path information, area boarder shall not leak the PPR-ID to the destination area.

3.3. PPR Management Plane Aspects

3.3.1. IGP Metric Independent Paths/Graphs

PPR allows a considerable simplification in the design and management of networks. In a best effort network the setting of the IGP metrics is a complex problem with competing constraints. A set of metrics that is optimal for traffic distribution under normal operation may not be optimal under conditions of failure of one or more of the network components. Nor is that choice of metrics necessarily best for operation under all IPFRR conditions. When SR is introduced to the network a further constraint on metrics is the need to limit the size of the SID stack/list. These problems further increase with the introduction of demanding technologies such as network slicing and deterministic networking.

Some mitigation occurs with the use of FlexAlgo [I-D.ietf-lsr-flex-algo] but fundamentally this is still an approach that is critically dependent on the per-flex-algo provisioning of different metrics on participating nodes, that operate in both the normal and the failure case.

PPR allows the network to simply introduce metric independent paths on a strategic or tactical basis. Being metric independent each PPR path operates ships-in-the-night with respect to all other paths. This means that the network management system can address network tuning on a case by case basis only needing to worry about the traffic matrix along the path rather than needing to deconvolve the impact of tuning a metric on the whole traffic matrix. In other words, PPR is a direct method of tuning the traffic rather than an the indirect method that metric tuning provides.

An example that makes this clear is the maximally redundant tree (MRT) approach to IPFRR. MRT requires the tuning of metrics to tune the paths, and a common algorithm for all nodes in the network. An equivalent solution can be introduced to the network by the insertion of a pair of PPR graphs by the network management system. Furthermore the topology of these graphs are independent of all other graphs, allowing the tuning and migration of the repair paths in the network management system.

Thus PPR allows the operator to focus on the desired traffic path of specific groups of packets independent of the desired path of the packets in all other paths.

3.3.2. Granular OAM

For some of the deployments as described in Section 2, the ability to collect certain statistics about PPR path usage, including how much traffic a PPR path carries and at what times from any node in the network is a critical requirement. Such statistics can be useful to account for the degree of usage of a path and provide additional operational insights, including usage patterns and trending information.

Traffic for certain PPRs may have more stringent requirement w.r.t accounting for critical SLAs (e.g. 5G non-eMBB slice) and should account for any link/node failures along the path. Optional per path attributes like "Packet Traffic Accounting" and "Traffic Statistics" instructs all the respective nodes along the path to provision the hardware and to account for the respective traffic statistics. Traffic accounting should be applied based on the PPR-ID. This capability allows a more granular and dynamic measurement of traffic statistics for only certain PPRs as needed.

As routing happens on the abstracted path identifier in the packet, no additional per packet instruction is needed for achieving the above functionality regardless of the data plane used in the network Section 3.1.

4. Preferred Path Loop Free Alternatives (pLFA)

PPR can be used as a method of providing IPFRR. Preferred Path Loop-Free Alternate (pLFA) allows the construction of arbitrary engineered backup paths pLFA and inherits the low packet overhead of PPR requiring a simple encapsulation and a single path identifier for any path of any complexity.

pLFA provides a superset of RSVP-TE repairs (complete with traffic engineering capability) and Topology Independent Loop-Free Alternates (TI-LFA) [I-D.ietf-rtgwg-segment-routing-ti-lfa]. However, unlike the TI-LFA approaches PPR is applicable to a more complete set of data planes (for example MPLS, both IPv4 and IPv6 and Ethernet) where it can provide a rich set of IPFRR capabilities ranging from simple best-effort repair calculated at the point of local repair (PLR) to full traffic engineered paths. For any repair path pLFA requires one encapsulation and one PPR-ID, regardless of the complexity and constraints of the path.

For a basic understanding of pLFA consider the case of a link repair shown in this example as shown in Figure 7, we assume that we have a path A-B-C-D that the packet must traverse. This may be a normal best effort path or a traffic engineered path.

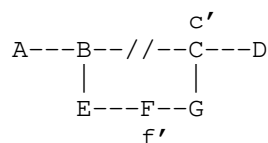


Figure 7

PPR is used to inject the repair path B->E->F->G->C into the network with a PPR-ID of c'. B is monitoring the health of link B->C, for example looking for loss-of-light, or using Bidirectional Forwarding Detection (BFD) [RFC5880]. When B detects a failure it encapsulates the packet to C by adding to the packet an encapsulation with a destination address set as the PPR-ID for c' and then sending the packet to E. At C the packet is decapsulated and sent to D. The path B->E->F->G->C may be a traffic engineered path or it may be a best effort path. This may of course be the post convergence path from B to C, as is used by TI-LFA However B may have at its disposal multiple paths to C with different properties for different traffic classes. In this case each path to be used would require its own PPR-ID (c', c'' etc.). Because pLFA only requires a single path identifier regardless of the complexity of the path is not necessary constrain the path to be a small number of loose source routed paths to protect against MTU or maximum SID count considerations.

pLFA supports the usual IPFRR features such as early release into Q-space, node repair, and shared risk link group support, LANs, ECMP and multi-homed prefixes. However the ability to apply repair graphs Section 3.2.5 is unique to pLFA. The use of graphs in IPFRR repair simplifies the construction of traffic engineered repair paths, and allows for the construction of arbitrary maximally redundant tree repair paths.

Of importance in any IPFRR strategy in a loosely routed network, including normal connectionless routing is the ability to support loop-free convergence. This problem is described in [RFC5715]. [I-D.ietf-rtgwg-segment-routing-ti-lfa] has proposed a mitigation technique for failures (noted above) and pLFA is able to support this. However a network supporting high reliability traffic may find mitigation insufficient. Also disruption can take place on network component inclusion (or repair/recovery) and TI-LFA is silent on this. A network using pLFA is compatible with all of the know loop-free convergence and loop mitigation approaches described in [RFC5715].

5. Traffic Engineering Attributes

In addition to determining the nodes to traverse, there may be other aspects that need to be set up for a path. Most notably, this concerns the allocation and reservation of resources along the path to help ensure the service levels, i.e. the Quality of Service that is delivered across the path, will be acceptable for the traffic routed across the path (critical in some deployments as listed in Section 2).

While SR allows packet steering on a specified path (for MPLS and IPv6 with SRH), it does not have any notion of QoS or resources reserved along the path. The determination of which resources to allocate and reserve on nodes across the path, like the determination of the path itself, can in many cases be made by a controller. Accordingly, PPR includes extensions that allow to manage those reservations, in addition to the path itself.

Key aspect of the solution concerns with specifying the resources to be reserved along the preferred path, through path attributes TLVs. Reservations are expressed in terms of required resources (bandwidth), traffic characteristics (burst size), and service level parameters (expected maximum latency at each hop) based on the capabilities of each node and link along the path. The second part of the solution is providing mechanism to indicate the status of the reservations requested i.e. if these have been honored by individual node/links in the path. This can be done by defining a new TLV/Sub-TLV in respective IGPs. Another aspect is additional node level TLVs

and extensions to IS-IS-TE [RFC8570] and OSPF-TE [RFC7471] to provide accounting/usage statistics that have to be maintained at each node per preferred path.

6. IANA Considerations

This document does not request any allocations from IANA.

7. Security Considerations

Advertisement of the additional information defined in this document introduces no new security concerns in IGP protocols. However, for extensions related to SR-MPLS and SRH data planes, those particular data plane security considerations does apply here.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

8.2. Informative References

- [I-D.ietf-isis-encapsulation-cap]
Xu, X., Decraene, B., Raszuk, R., Chunduri, U., Contreras, L. M., and L. Jalil, "Advertising Tunnelling Capability in IS-IS", draft-ietf-isis-encapsulation-cap-01 (work in progress), April 2017.
- [I-D.ietf-lsr-flex-algo]
Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", draft-ietf-lsr-flex-algo-19 (work in progress), April 2022.
- [I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", draft-ietf-rtgwg-segment-routing-ti-lfa-08 (work in progress), January 2022.

- [I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Network (VPN+) Services", draft-ietf-teas-enhanced-vpn-10 (work in progress), March 2022.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", RFC 5659, DOI 10.17487/RFC5659, October 2009, <<https://www.rfc-editor.org/info/rfc5659>>.
- [RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", RFC 5715, DOI 10.17487/RFC5715, January 2010, <<https://www.rfc-editor.org/info/rfc5715>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [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>>.
- [RFC7471] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", RFC 7471, DOI 10.17487/RFC7471, March 2015, <<https://www.rfc-editor.org/info/rfc7471>>.
- [RFC8300] Quinn, P., Ed., Elzur, U., Ed., and C. Pignataro, Ed., "Network Service Header (NSH)", RFC 8300, DOI 10.17487/RFC8300, January 2018, <<https://www.rfc-editor.org/info/rfc8300>>.
- [RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <<https://www.rfc-editor.org/info/rfc8402>>.
- [RFC8570] Ginsberg, L., Ed., Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 8570, DOI 10.17487/RFC8570, March 2019, <<https://www.rfc-editor.org/info/rfc8570>>.

- [RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.
- [RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <<https://www.rfc-editor.org/info/rfc8986>>.

Authors' Addresses

Stewart Bryant (editor)
University of Surrey 5GIC

Email: sb@stewartbryant.com

Uma Chunduri (editor)
Intel Corporation

Email: umac.ietf@gmail.com

Alexander Clemm
Futurewei

Email: ludwig@clemm.org

BFD Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2022

L. Han
M. Wang
China Mobile
F. Yang
Huawei Technologies
October 25, 2021

Signal Degrade Indication in BFD
draft-hwy-bfd-sdi-00

Abstract

To satisfy the requirements of signal degrade indication described in [I-D.yang-mpis-ps-sdi-sr], this document illustrates the extension of BFD protocol to support signal degrade indication.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 28, 2022.

Copyright Notice

Copyright (c) 2021 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Background	2
2. Terminology	2
3. Signal Degrade Overview	3
3.1. Signal Degrade Definition	3
3.2. Signal Degrade vs Packet Loss Rate	3
3.3. Use BFD to Support Signal Degrade Indication	4
3.4. Notification Spread in Network	4
4. BFD Extension to Indicate Signal Degrade	4
5. IANA Considerations	5
6. Security Considerations	5
7. References	5
7.1. Normative References	5
7.2. Informative References	6
Authors' Addresses	6

1. Background

Signal Degrade (SD) is categorized as one of triggers to bring survivability challenge to networks [RFC4428]. Not like the signal failure caused by failure of links or nodes, Signal Degrade (SD) is normally caused by fiber aging, fiber impairment, fiber pollution, optical module mismatch or WDM transmission error etc.

The detection and transmission of signal degrade is discussed in [I-D.zhl-mppls-tp-sd] and [I-D.yang-mppls-ps-sdi-sr]. When signal degrade is detected, it can be spread via control plane, forwarding plane, or management plane, or combination of any of them.

BFD [RFC5880] and SBFD [RFC7880] are widely used as the failure notification in networks due to the characteristics of simplicity and efficiency. BFD also provides good opportunity to indicate signal degrade by reflecting it in BFD state changes. This document extends the BFD protocol to carry signal degrade indication in networks.

2. Terminology

SD: Signal Degrade

BER: Bit Error Rate

MIP: Maintenance Entity Group Intermediate Point

PLR: Packet Loss Rate

FEC: Forwarding Error Correction

SLA: Service Level Agreement

BFD: Bidirectional Forwarding Detection

SBFD: Seamless BFD

OAM: Operation, Administration and Maintenance

3. Signal Degrade Overview

3.1. Signal Degrade Definition

In [IEEE 802.3-2018], Bit Error Rate (BER) is defined as the ratio of the number of bits received in error to the total number of bits received. It is one of parameters to indicate quality of physical links. Depending on the Forwarding Error Correction (FEC) capability of PHYs, BER can be classified into pre-FEC BER and post-FEC BER. The pre-FEC BER value acquired from PHY on receiving port indicates the on wire BER value of physical link. This value can also be measured via external test instruments. Generally speaking, BER specifically refers to pre-FEC BER. If FEC capability is unavailable for some legacy PHYs, it is meaningless to differentiate pre-FEC and post-FEC BER values.

Signal degrade can be detected based on the physical bit error statistic on port level, no matter whether the PHY is with or without Forwarding Error Correction. Port level statistic is an intuitive approach to be best understood in the equipment and network systems. In practice, flexible configuration of the watermark to trigger the indication of signal degrade is also preferred.

3.2. Signal Degrade vs Packet Loss Rate

In packet switched network, the measurement of physical link is based on the unit of packet, resulting in either no packet loss or a number of packet loss to indicate the status of link. Although PHYs are defined in [IEEE 802.3-2018], vendors may have different implementations to deal with the error bits when equipment detects them. Moreover, bit is a fix unit, but packet has variable length. Several error bits can lead to one packet loss, or multiple packets' loss. There is no uniform approach to calculate pre-FEC BER into

PLR. It means there is no parameter directly indicated the status of physical links in packet switched network.

3.3. Use BFD to Support Signal Degrade Indication

For the network where BFD is used to provide the fast failure detection, the minimal detection interval e.g. 3.3ms actually leaves a huge gap of data packets between two consecutive BFD packets when the line rate packets are transmitted over high speed Ethernet link. Take an example of 10Gbps link transmitting the packets with length of 192 bytes to calculate, more than twenty thousand packets are transmitted within 3.3ms. Note that the criteria to announce a failure of BFD based on three consecutive BFD packet loss. It may not be accurate to rely on BFD to detect and trigger the protection mechanism if there is signal degrade on the physical link.

3.4. Notification Spread in Network

In current packet switched networks, the error bit information like BER is only obtained and processed locally on each node. There is no indication or advertisement of the errors or its indications of physical links. It should be possible to spread this information via control plane, management plane or even data plane to suit for different needs. Especially, if the signal degrade of the link could be transmitted in data plane and aware by any other nodes, local repair or end-to-end path protection could be performed even more efficiently. Previous work proposed in [I-D.rkhd-mpls-tp-sd], [I-D.zhl-mpls-tp-sd] and [I-D.zhang-ccamp-rsvpte-ber-measure] give the examples of protocol extensions to support SD transmission for further network convergence behaviors. With the emerge of telemetry, it is also possible to collect and report this information more frequently to SDN controller to facilitate the network operation and management.

4. BFD Extension to Indicate Signal Degrade

The Diagnostic code in BFD specifies the local system's reason for the last change in session state. The definition of the Values is specified in Section 4.1 of [RFC5880].

In this document, reserved values from 9 to 31 are requested to IANA to support the signal degrade indication and removal.

(preamble)

0										1										2										3									
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1								
Vers										Sta P F C A D M										Detect Mult										Length									
										My Discriminator																													
										Your Discriminator																													
										Desired Min TX Interval																													
										Required Min RX Interval																													
										Required Min Echo RX Interval																													
										Authentication (optional)																													

BFD Packet Format

5. IANA Considerations

The document requires the definition of the new indication and removal of the signal degrade indication in BFD Value code.

6. Security Considerations

TBD

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [RFC7880] Pignataro, C., Ward, D., Akiya, N., Bhatia, M., and S. Pallagatti, "Seamless Bidirectional Forwarding Detection (S-BFD)", RFC 7880, DOI 10.17487/RFC7880, July 2016, <<https://www.rfc-editor.org/info/rfc7880>>.

7.2. Informative References

- [I-D.rkhd-mpls-tp-sd]
Ram, R., Cohn, D., Daikoku, M., Yuxia, M., Jian, Y., and A. D'Alessandro, "SD detection and protection triggering in MPLS-TP", draft-rkhd-mpls-tp-sd-03 (work in progress), May 2011.
- [I-D.yang-mpls-ps-sdi-sr]
Yang, F., Han, L., and J. Zhao, "Problem Statement of Signal Degrade Indication for SR over MPLS", draft-yang-mpls-ps-sdi-sr-01 (work in progress), November 2020.
- [I-D.zhang-ccamp-rsvpte-ber-measure]
Li, Z., Zhang, L., and G. Yang, "RSVP-TE Extensions for Bit Error Rate (BER) Measurement", draft-zhang-ccamp-rsvpte-ber-measure-02 (work in progress), July 2014.
- [I-D.zhl-mpls-tp-sd]
Haiyan, Z., Jia, H., and H. Li, "SD-Triggered Protection Switching in MPLS-TP", draft-zhl-mpls-tp-sd-03 (work in progress), October 2010.
- [RFC4428] Papadimitriou, D., Ed. and E. Mannie, Ed., "Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanisms (including Protection and Restoration)", RFC 4428, DOI 10.17487/RFC4428, March 2006, <<https://www.rfc-editor.org/info/rfc4428>>.
- [RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, DOI 10.17487/RFC5586, June 2009, <<https://www.rfc-editor.org/info/rfc5586>>.
- [RFC6372] Sprecher, N., Ed. and A. Farrel, Ed., "MPLS Transport Profile (MPLS-TP) Survivability Framework", RFC 6372, DOI 10.17487/RFC6372, September 2011, <<https://www.rfc-editor.org/info/rfc6372>>.

Authors' Addresses

Liuyan Han
China Mobile
No.32 Xuanwumen west street
Beijing 100053
China

Email: hanliuyan@chinamobile.com

Minxue Wang
China Mobile
No.32 Xuanwumen west street
Beijing 100053
China

Email: wangminxue@chinamobile.com

Fan Yang
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: shirley.yangfan@huawei.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: October 20, 2022

Z. Hu
Huawei
H. Chen
Futurewei
H. Chen
China Telecom
P. Wu
Huawei
M. Toy
Verizon
C. Cao
T. He
China Unicom
L. Liu
Fujitsu
X. Liu
Volta Networks
April 18, 2022

SRv6 Path Egress Protection
draft-ietf-rtgwg-srv6-egress-protection-05

Abstract

This document describes protocol extensions for protecting the egress node of a Segment Routing for IPv6 (SRv6) path or tunnel.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 20, 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Terminologies	3
3. SR Path Egress Protection	4
3.1. Mechanism	4
3.2. Example	6
4. Extensions to IGP for Egress Protection	8
4.1. Extensions to IS-IS	9
4.2. Extensions to OSPF	11
5. Security Considerations	13
6. IANA Considerations	13
6.1. SRv6 Endpoint Behaviors	13
6.2. IS-IS	14
6.3. OSPFv3	14
7. Acknowledgements	15
8. References	15
8.1. Normative References	15
8.2. Informative References	16
Authors' Addresses	17

1. Introduction

The fast protection of a transit node of a Segment Routing (SR) path or tunnel is described in [I-D.ietf-rtgwg-segment-routing-ti-lfa] and [I-D.hu-spring-segment-routing-proxy-forwarding]. [RFC8400] specifies the fast protection of egress node(s) of an MPLS TE LSP tunnel including P2P TE LSP tunnel and P2MP TE LSP tunnel in details. However, these documents do not discuss the fast protection of the egress node of a Segment Routing for IPv6 (SRv6) path or tunnel.

This document fills that void and presents protocol extensions for the fast protection of the egress node of an SRv6 path or tunnel. Egress node and egress, fast protection and protection as well as SRv6 path and SRv6 tunnel will be used exchangeably below.

There are a number of topics related to the egress protection, which include the detection of egress node failure, the relation between egress protection and global repair, and so on. These are discussed in details in [RFC8679].

2. Terminologies

The following terminologies are used in this document.

SR: Segment Routing

SRv6: SR for IPv6

SRH: Segment Routing Header

SID: Segment Identifier

LSA: Link State Advertisement in OSPF

LSP: Label Switched Path in MPLS or Link State Protocol PDU in IS-IS

PDU: Protocol Data Unit

LS: Link State, which is LSA in OSPF or LSP in IS-IS

TE: Traffic Engineering

SA: Source Address

DA: Destination Address

P2MP: Point-to-MultiPoint

P2P: Point-to-Point

CE: Customer Edge

PE: Provider Edge

LFA: Loop-Free Alternate

TI-LFA: Topology Independent LFA

BFD: Bidirectional Forwarding Detection

VPN: Virtual Private Network

L3VPN: Layer 3 VPN

VRF: Virtual Routing and Forwarding

FIB: Forwarding Information Base

PLR: Point of Local Repair

BGP: Border Gateway Protocol

IGP: Interior Gateway Protocol

OSPF: Open Shortest Path First

IS-IS: Intermediate System to Intermediate System

3. SR Path Egress Protection

This section describes the mechanism of SR path egress protection and illustrates it through an example.

3.1. Mechanism

Figure 1 is used to explain the mechanism of SR path egress node protection.

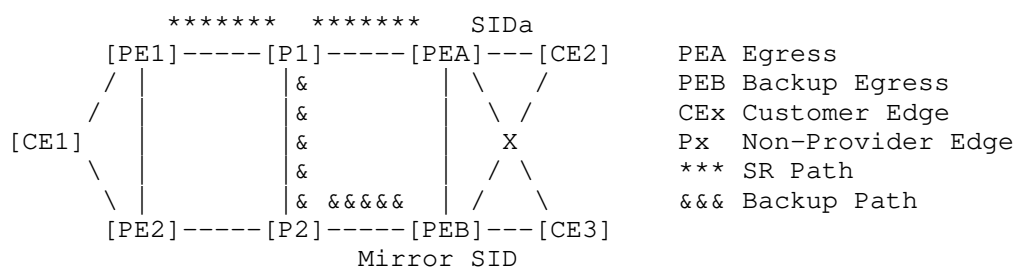


Figure 1: PEB Protects Egress PEA of SR Path

Where node PEA is the egress of the SR path from PE1 to PEA, and has SIDA which is the active segment in the packet from the SR path at PEA. Node PEB is the backup egress (or say protector) to provide the protection for egress (or say primary egress) PEA. Node P1 is the direct previous hop of egress PEA and acts as PLR to support the protection for PEA.

When PEB is selected as a backup egress to protect the egress PEA, a Mirror SID (refer to Section 5.1 of [RFC8402]) is configured on PEB to protect PEA. PEB advertises this information through IGP, which includes the Mirror SID and the egress PEA. The information is represented by <PEB, PEA, Mirror SID>, which indicates that PEB protects PEA with Mirror SID.

After PEA receives the information <PEB, PEA, Mirror SID>, it may send the forwarding behavior of the SIDA at PEA to PEB with the Mirror SID using some protocols such as BGP if PEB can not obtain this behavior from other approaches if PEB wants to protect SIDA of PEA. How to send the forwarding behavior of the SIDA to PEB is out scope of this document.

When PEB gets the forwarding behavior of the SIDA of PEA from PEA or other means, it adds a forwarding entry for the SIDA according to the behavior into the forwarding table for node PEA. This table is identified by the Mirror SID, which indicates node PEA's context. Using the forwarding entry for SIDA in this table, a packet with SIDA will be transmitted by PEB to the same destination as it is transmitted by PEA. For example, assume that the packet with SIDA is transmitted by PEA to CE2 through the forwarding behavior of the SIDA in PEA. The packet will be transmitted by PEB to the same CE2 through looking up the table identified by the Mirror SID.

After P1 as PLR receives the information <PEB, PEA, Mirror SID> and knows that PEB wants to protect SIDA of PEA, it computes a shortest path to PEB. A Repair List RL is obtained based on the path. It is one of the followings:

- o RL = <Mirror SID> if the path does not go through PEA; or
- o RL = <S1, ..., Sn, Mirror SID> if the path goes through PEA, where <S1, ..., Sn> is the TI-LFA Repair List to PEB computed by P1.

When PEA fails, P1 as PLR sends the packet with SIDA carried by the SR path to PEB, but encapsulates the packet before sending it by executing H.Encaps with the Repair List RL and a Source Address T.

Suppose that the packet received by P1 is represented by Pkt = (S, SIDA)Pkt0, where SA = S and DA = SIDA, and Pkt0 is the rest of the packet.

The execution of H.Encaps pushes an IPv6 header to Pkt and sets some fields in the outer and inner IPv6 header to produce an encapsulated packet Pkt'. Pkt' will be one of the followings:

- o Pkt' = (T, Mirror SID) (S, SIDA)Pkt0 if RL = <Mirror SID>; or

- o $Pkt' = (T, S1)(Mirror\ SID, Sn, \dots, S1; SL=n) (S, SIDA)Pkt0$ if $RL = \langle S1, \dots, Sn, Mirror\ SID \rangle$.

When PEB receives the re-routed packet, which is $(T, Mirror\ SID) (S, SIDA)Pkt0$, it decapsulates the packet and forwards the decapsulated packet using the FIB table Tm identified by the Mirror SID as a variant of End.DT6 SID. The Mirror SID is called End.M.

It obtains the Mirror SID in the outer IPv6 header of the packet, removes this outer IPv6 header with all its extension headers, and then processes the inner IPv6 packet (i.e., $(S, SIDA)Pkt0$, the packet without the outer IPv6 header). PEB finds the FIB table Tm for node PEA using the Mirror SID as the context ID, and submits the packet to this FIB table lookup and transmission to the same destination as PEA does.

The behavior of Mirror SID (End.M for short) is a variant of the End.DT6 behavior (refer to Section 4.6 of [RFC8986]). The End.M SID MUST be the last segment in an SR path, and a SID instance is associated with an IPv6 FIB table Tm .

When processing the Upper-Layer header of a packet matching a FIB entry locally instantiated as an End.M SID, N does the following:

```

S01. If (Upper-Layer header type == 41(IPv6) ) {
S02.   Remove the outer IPv6 header with all its extension headers
S03.   Set the packet's associated FIB table to  $Tm$ 
S04.   Submit the packet to the egress IPv6 FIB lookup for
       transmission to the new destination
S05. } Else {
S06.   Process as per Section 4.1.1 of RFC8986
S07. }
```

3.2. Example

Figure 2 shows an example of protecting egress PE3 of a SR path, which is from ingress PE1 to egress PE3.

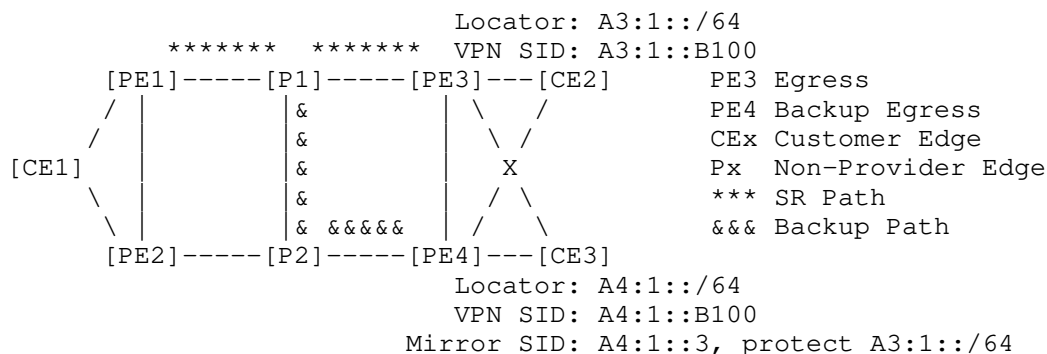


Figure 2: PE4 Protects Egress PE3 of SR Path

Where node P1's pre-computed backup path for PE3 is from P1 to PE4 via P2. In normal operations, after receiving a packet with destination PE3, P1 forwards the packet to PE3 according to its FIB. When PE3 receives the packet, it sends the packet to CE2.

When PE3 fails, P1 as PLR detects the failure through using a failure detection mechanism such as BFD and forwards the packet to PE4 via the backup path. When PE4 receives the packet, it sends the packet to the same CE2.

In Figure 2, Both CE2 and CE3 are dual home to PE3 and PE4. PE3 has a locator A3:1::/64 and a VPN SID A3:1::B100. PE4 has a locator A4:1::/64 and VPN SID A4:1::B100. A Mirror SID A4:1::3 is configured on PE4 for protecting PE3 with locator A3:1::/64.

After the configuration, PE4 advertises this information through an IGP LS (i.e., LSA in OSPF or LSP in IS-IS), which includes PE3's locator and Mirror SID A4:1::3. Every node in the SR domain will receive this IGP LS, which indicates that PE4 wants to protect PE3's locator with Mirror SID A4:1::3.

When PE4 (e.g., BGP on PE4) receives a prefix whose VPN SID belongs to PE3 that is protected by PE4 through Mirror SID A4:1::3, it finds PE4's VPN SID corresponding to PE3's VPN SID. For example, local PE4 has Prefix 1.1.1.1 with VPN SID A4:1::B100, when PE4 receives prefix 1.1.1.1 with remote PE3's VPN SID A3:1::B100, it knows that they are for the same VPN.

The forwarding behaviors for these two VPN SIDs are the same from function's point of view. If the behavior for PE3's VPN SID in PE3 forwards the packet with it to CE2, then the behavior for PE4's VPN SID in PE4 forwards the packet to the same CE2; and vice versa. PE4 creates a forwarding entry for PE3's VPN SID A3:1::B100 in the FIB

table identified by Mirror SID A4:1::3 according to the forwarding behavior for PE4's VPN SID A4:1::B100.

Node P1's pre-computed backup path for destination PE3's locator is from P1 to PE4 having mirror SID A4:1::3. When P1 receives a packet destined to PE3's VPN SID A3:1::B100, in normal operations, it forwards the packet with source A1:1:: and destination PE3's VPN SID A3:1::B100 according to the FIB using the destination PE3's VPN SID A3:1::B100.

When PE3 fails, P1 as PLR sends the packet to PE4 via the backup path pre-computed. P1 encapsulates the packet using H.Encaps before sending it to PE4.

Suppose that the packet received by P1 is represented by Pkt = (SA = A1:1::, DA = A3:1::B100)Pkt0, where DA = A3:1::B100 is PE3's VPN SID, and Pkt0 is the rest of the packet. The encapsulated packet Pkt' will be one of the followings:

- o Pkt' = (T, Mirror SID A4:1::3) (A1:1::, A3:1::B100)Pkt0 if backup path not via PE3; or (otherwise)
- o Pkt' = (T, S1) (Mirror SID A4:1::3, Sn, ..., S1; SL=n) (A1:1::, A3:1::B100)Pkt0.

where T is a Source Address, <S1, ..., Sn> is the TI-LFA Repair List to PE4 computed by P1 when the backup path to PE4 goes through PE3.

When PE4 receives the re-routed packet, it decapsulates the packet and forwards the decapsulated packet by executing End.DT6 behavior for an End.DT6 SID instance. The SID instance is End.M, the Mirror SID that is associated with the IPv6 FIB table for PE3. The packet received by PE4 is (T, Mirror SID A4:1::3) (A1:1::, PE3's VPN SID A3:1::B100)Pkt0.

PE4 obtains Mirror SID A4:1::3 in the outer IPv6 header of the packet, removes this outer IPv6 header, and then processes the inner IPv6 packet (A1:1::, A3:1::B100)Pkt0. It finds the FIB table for PE3 using Mirror SID A4:1::3 as the context ID, gets the forwarding entry for PE3's VPN SID A3:1::B100 from the table, and forwards the packet to CE2 using the entry.

4. Extensions to IGP for Egress Protection

This section describes extensions to IS-IS and OSPF for advertising the information about SRv6 path egress protection.

4.1. Extensions to IS-IS

A new sub-TLV, called IS-IS SRv6 Mirror SID sub-TLV, is defined. It is used in the SRv6 Locator TLV defined in [I-D.ietf-lsr-isis-srv6-extensions] to advertise SRv6 Mirror SID and the locators of the node to be protected. The SRv6 Mirror SID inherit the topology/algorithm from the parent locator. The format of the sub-TLV is illustrated below.

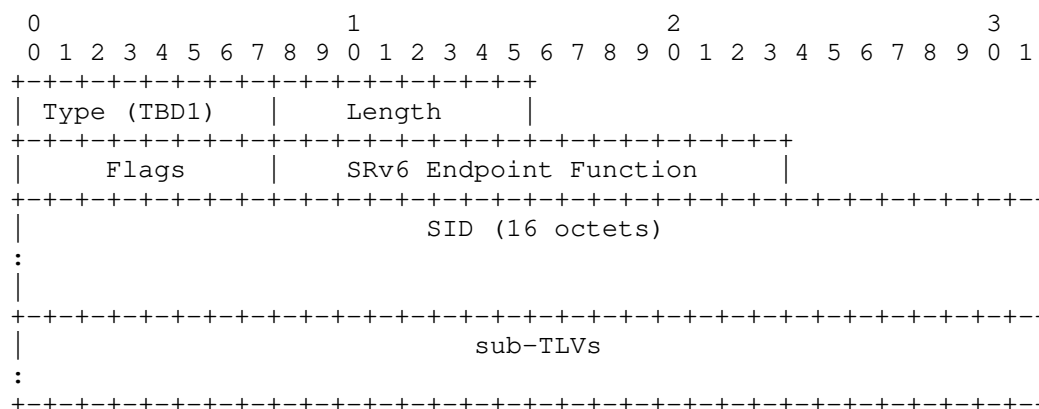


Figure 3: IS-IS SRv6 Mirror SID sub-TLV

Type: TBD1 (suggested value 8) is to be assigned by IANA.

Length: variable.

Flags: 1 octet. No flags are currently defined.

SRv6 Endpoint Function: 2 octets. It contains the endpoint function 74 for Mirror SID.

SID: 16 octets. This field contains the SRv6 Mirror SID to be advertised.

Two sub-TLVs are defined. One is the protected locators sub-TLV, and the other is the protected SIDs sub-TLV.

A protected locators sub-TLV is used to carry the Locators to be protected by the SRv6 mirror SID. It has the following format.

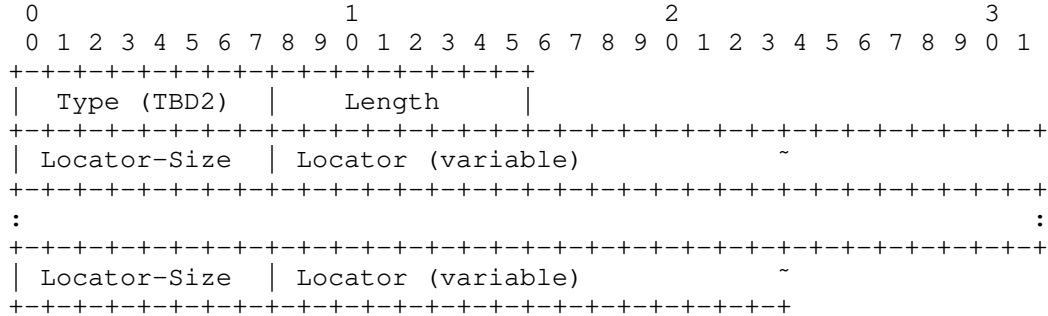


Figure 4: IS-IS Protected Locators sub-TLV

Type: TBD2 (suggested value 1) is to be assigned by IANA.

Length: variable.

Locator-Size: 1 octet. Number of bits (1 - 128) in the Locator field.

Locator: 1-16 octets. This field encodes an SRv6 Locator to be protected by the SRv6 mirror SID. The Locator is encoded in the minimal number of octets for the given number of bits.

A protected SIDs sub-TLV is used to carry the SIDs to be protected by the SRv6 Mirror SID. It has the following format.

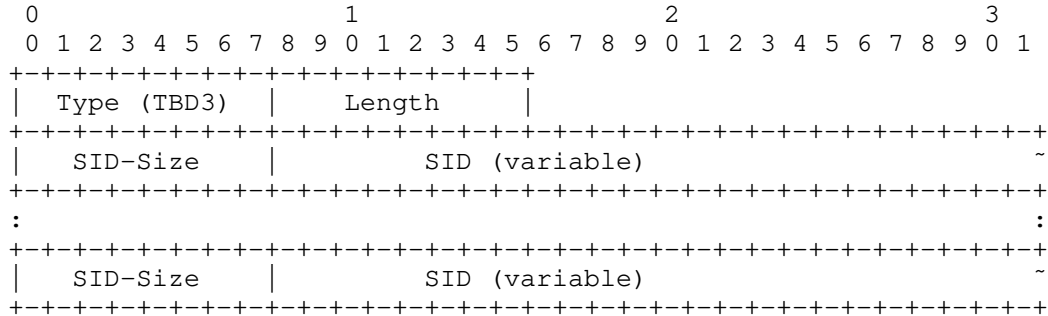


Figure 5: IS-IS Protected SIDs sub-TLV

Type: TBD3 (suggested value 2) is to be assigned by IANA.

Length: variable.

SID-Size: 1 octet. Number of bits in the SID field. It is from 1 to 128. When it is less than 128, the SID field is a locator. When it is 128, the SID field is an SRv6 SID.

SID: 1-16 octets. This field encodes an SRv6 SID or locator to be protected. The SID/locator is encoded in the minimal number of octets for the given number of bits. Trailing bits MUST be set to zero and ignored when received.

When node B advertises that B wants to protect node A's locators with a Mirror SID through an LSP, the LSP contains an IS-IS SRv6 Mirror SID sub-TLV, which includes the Mirror SID and the node A's locators in an IS-IS Protected locators sub-TLV. If B wants to protect just a specific set of SIDs of node A, the Mirror SID sub-TLV includes these SIDs in an IS-IS Protected SIDs sub-TLV.

4.2. Extensions to OSPF

Similarly, a new sub-TLV, called OSPF Mirror SID sub-TLV, is defined. It is used to advertise SRv6 Mirror SID and the locators of the node to be protected. Its format is illustrated below.

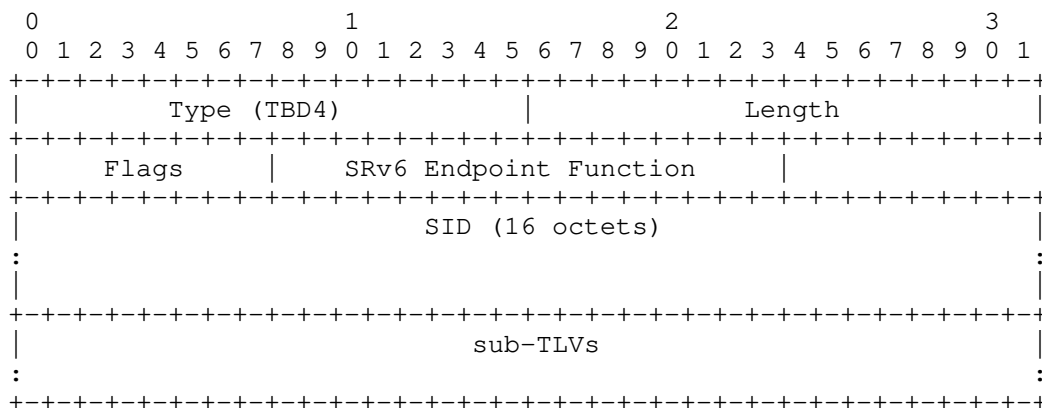


Figure 6: OSPF SRv6 Mirror SID sub-TLV

Type: TBD4 (suggested value 8) is to be assigned by IANA.

Length: variable.

Flags: 1 octet. No flags are currently defined.

SRv6 Endpoint Function: 2 octets. It contains the endpoint function 74 for End.M SID.

SID: 16 octets. This field contains the SRv6 Mirror SID to be advertised.

Two sub-TLVs are defined. One is the protected locators sub-TLV, and the other is the protected SIDs sub-TLV.

A protected locators sub-TLV is used to carry the locators of the node to be protected by the SRv6 Mirror SID. It has the following format.

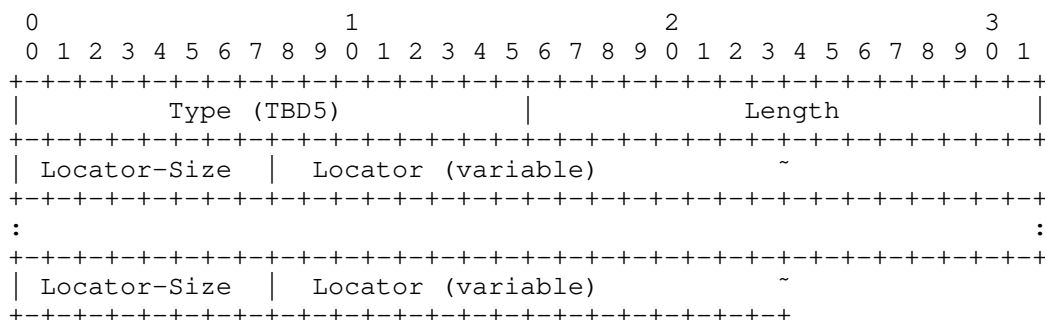


Figure 7: OSPF Protected Locators sub-TLV

Type: TBD5 (suggested value 1) is to be assigned by IANA.

Length: variable.

Locator-Size: 1 octet. Number of bits (1 - 128) in the Locator field.

Locator: 1-16 octets. This field encodes an SRv6 Locator to be protected by the SRv6 mirror SID. The Locator is encoded in the minimal number of octets for the given number of bits.

A protected SIDs sub-TLV is used to carry the SIDs to be protected by the SRv6 Mirror SID. It has the following format.

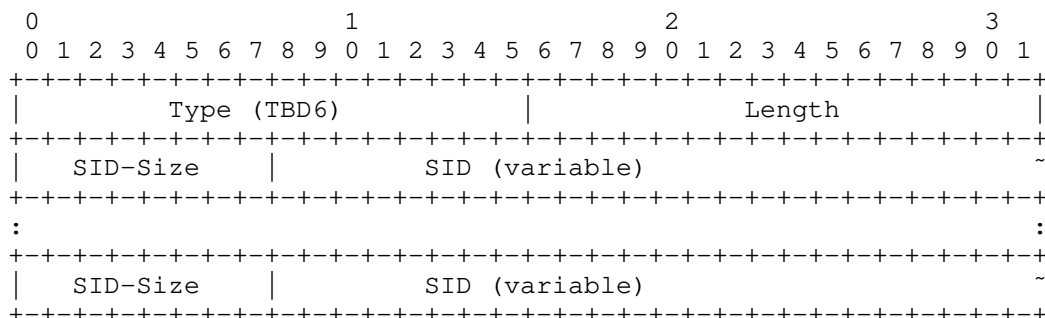


Figure 8: OSPF Protected SIDs sub-TLV

Type: TBD6 (suggested value 2) is to be assigned by IANA.

Length: variable.

SID-Size: 1 octet. Number of bits in the SID field. It is from 1 to 128. When it is less than 128, the SID field is a locator. When it is 128, the SID field is an SRv6 SID.

SID: 1-16 octets. This field encodes an SRv6 SID or locator to be protected. The SID/locator is encoded in the minimal number of octets for the given number of bits. Trailing bits MUST be set to zero and ignored when received.

5. Security Considerations

The security about the egress protection is described in details in [RFC8679]. The extensions to OSPF and IS-IS described in this document for SRv6 path egress protection should not cause extra security issues.

6. IANA Considerations

6.1. SRv6 Endpoint Behaviors

Under sub-registry "SRv6 Endpoint Behaviors" [RFC8986], IANA has assigned the following for End.M Endpoint Behavior:

Value	Hex	Endpoint behavior	Reference
74	0x004A	End.M (Mirror SID)	This document

6.2. IS-IS

Under "Sub-TLVs for TLVs 27, 135, 235, 236 and 237 registry" [I-D.ietf-lsr-isis-srv6-extensions], IANA is requested to add the following new Sub-TLV:

Sub-TLV Type	Sub-TLV Name	Reference
8	SRv6 Mirror SID Sub-TLV	This document

IANA is requested to create and maintain a new registry for sub-sub-TLVs of the SRv6 Mirror SID Sub-TLV. The suggested registry name is

- o Sub-Sub-TLVs for SRv6 Mirror SID Sub-TLV

Initial values for the registry are given below. The future assignments are to be made through IETF Review [RFC5226].

Value	Sub-Sub-TLV Name	Definition
0	Reserved	
1	Protected Locators Sub-Sub-TLV	This Document
2	Protected SIDs Sub-Sub-TLV	
3-255	Unassigned	

6.3. OSPFv3

Under registry "OSPFv3 Locator LSA Sub-TLVs" [I-D.ietf-lsr-ospfv3-srv6-extensions], IANA is requested to assign the following new Sub-TLV:

Sub-TLV Type	Sub-TLV Name	Reference
8	SRv6 Mirror SID Sub-TLV	This document

IANA is requested to create and maintain a new registry for sub-sub-TLVs of the SRv6 Mirror SID Sub-TLV. The suggested registry name is

- o Sub-Sub-TLVs for SRv6 Mirror SID Sub-TLV

Initial values for the registry are given below. The future assignments are to be made through IETF Review [RFC5226].

Value	Sub-Sub-TLV Name	Definition
-----	-----	-----
0	Reserved	
1	Protected Locators Sub-Sub-TLV	This Document
2	Protected SIDs Sub-Sub-TLV	
3-65535	Unassigned	

7. Acknowledgements

The authors would like to thank Peter Psenak, Yimin Shen, Zhenqiang Li, Alexander Vainshtein, Greg Mirsky, Bruno Decraene, Jeff Tantsura, Chris Bowers and Ketan Talaulikar for their comments to this work.

8. References

8.1. Normative References

- [I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Segment Routing over IPv6 Dataplane", draft-ietf-lsr-isis-srv6-extensions-18 (work in progress), October 2021.
- [I-D.ietf-lsr-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", draft-ietf-lsr-ospfv3-srv6-extensions-03 (work in progress), November 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>>.
- [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, DOI 10.17487/RFC7356, September 2014, <<https://www.rfc-editor.org/info/rfc7356>>.
- [RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <<https://www.rfc-editor.org/info/rfc7490>>.
- [RFC8400] Chen, H., Liu, A., Saad, T., Xu, F., and L. Huang, "Extensions to RSVP-TE for Label Switched Path (LSP) Egress Protection", RFC 8400, DOI 10.17487/RFC8400, June 2018, <<https://www.rfc-editor.org/info/rfc8400>>.

- [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>>.
- [RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <<https://www.rfc-editor.org/info/rfc8665>>.
- [RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <<https://www.rfc-editor.org/info/rfc8667>>.
- [RFC8679] Shen, Y., Jeganathan, M., Decraene, B., Gredler, H., Michel, C., and H. Chen, "MPLS Egress Protection Framework", RFC 8679, DOI 10.17487/RFC8679, December 2019, <<https://www.rfc-editor.org/info/rfc8679>>.
- [RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <<https://www.rfc-editor.org/info/rfc8986>>.

8.2. Informative References

- [I-D.hu-spring-segment-routing-proxy-forwarding]
Hu, Z., Chen, H., Yao, J., Bowers, C., Yongqing, and Yisong, "SR-TE Path Midpoint Restoration", draft-hu-spring-segment-routing-proxy-forwarding-19 (work in progress), April 2022.
- [I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", draft-ietf-rtgwg-segment-routing-ti-lfa-08 (work in progress), January 2022.
- [I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-22 (work in progress), March 2022.

- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, DOI 10.17487/RFC5226, May 2008, <<https://www.rfc-editor.org/info/rfc5226>>.
- [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>>.

Authors' Addresses

Zhibo Hu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: huzhibo@huawei.com

Huaimo Chen
Futurewei
Boston, MA
USA

Email: Huaimo.chen@futurewei.com

Huanan Chen
China Telecom
109, West Zhongshan Road, Tianhe District
Guangzhou 510000
China

Email: chenhn8.gd@chinatelecom.cn

Peng Wu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: baggio.wupeng@huawei.com

Mehmet Toy
Verizon
USA

Email: mehmet.toy@verizon.com

Chang Cao
China Unicom
Beijing China

Email: caoc15@chinaunicom.cn

Tao He
China Unicom
Beijing China

Email: het21@chinaunicom.cn

Lei Liu
Fujitsu
USA

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks
McLean, VA
USA

Email: xufeng.liu.ietf@gmail.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 17 August 2022

L. Han, Ed.
R. Li
A. Retana
Futurewei Technologies, Inc.
M. Chen
L. Su
China Mobile
N. Wang
University of Surrey
13 February 2022

Problems and Requirements of Satellite Constellation for Internet
draft-lhan-problems-requirements-satellite-net-02

Abstract

This document presents the detailed analysis about the problems and requirements of satellite constellation used for Internet. It starts from the satellite orbit basics, coverage calculation, then it estimates the time constraints for the communications between satellite and ground-station, also between satellites. How to use satellite constellation for Internet is discussed in detail including the satellite relay and satellite networking. The problems and requirements of using traditional network technology for satellite network integrating with Internet are finally outlined.

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 17 August 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	3
2. Terminology	3
3. Overview	5
4. Basics of Satellite Constellation	6
4.1. Satellite Orbit	6
4.2. Coverage of LEO and VLEO Satellites and Minimum Number Required	6
4.3. Real Deployment of LEO and VLEO for Satellite Network . .	9
5. Communications for Satellite Constellation	10
5.1. Dynamic Ground-station-Satellite Communication	11
5.2. Dynamic Inter-satellite Communication	12
5.2.1. Inter-satellite Communication Overview	12
5.2.2. Satellites on Adjacent Orbit Planes with Same Altitude	15
5.2.3. Satellites on Adjacent Orbit Planes with Different Altitude	17
6. Use Satellite Network for Internet	19
7. Problems and Requirements for Satellite Constellation for Internet	22
7.1. Common Problems and Requirements	22
7.2. Satellite Relay	25
7.2.1. One Satellite Relay	25
7.2.2. Multiple Satellite Relay	26
7.3. Satellite Networking	28
7.3.1. L2 or L3 network	28
7.3.2. Inter-satellite-Link Lifetime	28
7.3.3. Problems for Traditional Routing Technologies	29
8. IANA Considerations	33
9. Contributors	33
10. Acknowledgements	33
11. References	33
11.1. Normative References	33
11.2. Informative References	34
Appendix A. Change Log	36
Authors' Addresses	36

1. Introduction

Satellite constellation for Internet is emerging. Even there is no constellation network established completely yet at the time of the publishing of the draft (June 2021), some basic internet service has been provided and has demonstrated competitive quality to traditional broadband service.

This memo will analyze the challenges for satellite network used in Internet by traditional routing and switching technologies. It is based on the analysis of the dynamic characters of both ground-station-to-satellite and inter-satellite communications and its impact to satellite constellation networking.

The memo also provides visions for the future solution, such as in routing and forwarding.

The memo focuses on the topics about how the satellite network can work with Internet. It does not focus on physical layer technologies (wireless, spectrum, laser, mobility, etc.) for satellite communication.

2. Terminology

LEO	Low Earth Orbit with the altitude from 180 km to 2000 km.
VLEO	Very Low Earth Orbit with the altitude below 450 km
MEO	Medium Earth Orbit with the altitude from 2000 km to 35786 km
GEO	Geosynchronous orbit with the altitude 35786 km
GSO	Geosynchronous satellite on GEO
ISL	Inter Satellite Link
ISLL	Inter Satellite Laser Link
EIRP	Effective isotropic radiated power
P2MP	Point to Multiple Points
GS	Ground Station, a device on ground connecting the

satellite. In the document, GS will hypothetically provide L2 and/or L3 functionality in addition to process/send/receive radio wave. It might be different as the reality that the device to process/send/receive radio wave and the device to provide L2 and/or L3 functionality could be separated.

SGS	Source ground station. For a specified flow, a ground station that will send data to a satellite through its uplink.
DGS	Destination ground station. For a specified flow, a ground station that is connected to a local network or Internet, it will receive data from a satellite through its downlink and then forward to a local network or Internet.
PGW	Packet Gateway
UPF	User Packet Function
PE router	Provider Edge router
CE router	Customer Edge router
P router	Provider router
LSA	Link-state advertisement
LSP	Link-State PDUs
L1	Layer 1, or Physical Layer in OSI model [OSI-Model]
L2	Layer 2, or Data Link Layer in OSI model [OSI-Model]
L3	Layer 3, or Network Layer in OSI model [OSI-Model], it is also called IP layer in TCP/IP model
BGP	Border Gateway Protocol [RFC4271]
eBGP	External Border Gateway Protocol, two BGP peers have different Autonomous Number
iBGP	Internal Border Gateway Protocol, two BGP peers have same Autonomous Number

IGP Interior gateway protocol, examples of IGPs include Open Shortest Path First (OSPF [RFC2328]), Routing Information Protocol (RIP [RFC2453]), Intermediate System to Intermediate System (IS-IS [RFC7142]) and Enhanced Interior Gateway Routing Protocol (EIGRP [RFC7868]).

3. Overview

The traditional satellite communication system is composed of few GSO and ground stations. For this system, each GSO can cover 42% Earth's surface [GEO-Coverage], so as few as three GSO can provide the global coverage theoretically. With so huge coverage, GSO only needs to amplify signals received from uplink of one ground station and relay to the downlink of another ground station. There is no inter-satellite communications needed. Also, since the GSO is stationary to the ground station, there is no mobility issue involved.

Recently, more and more LEO and VLEO satellites have been launched, they attract attentions due to their advantages over GSO and MEO in terms of higher bandwidth, lower cost in satellite, launching, ground station, etc. Some organizations [ITU-6G][Surrey-6G][Nttddocomo-6G] have proposed the non-terrestrial network using LEO, VLEO as important parts for 6G to extend the coverage of Internet. SpaceX has started to build the satellite constellation called StarLink that will deploy over 10 thousand LEO and VLEO satellites finally [StarLink]. China also started to request the spectrum from ITU to establish a constellation that has 12992 satellites [China-constellation]. European Space Agency (ESA) has proposed "Fiber in the sky" initiative to connect satellites with fiber network on Earth [ESA-HyDRON].

When satellites on MEO, LEO and VLEO are deployed, the communication problem becomes more complicated than for GSO. This is because the altitude of MEO/LEO/VLEO satellites are much lower. As a result, the coverage of each satellite is much smaller than for GSO, and the satellite is not relatively stationary to the ground. This will lead to:

1. More satellites than GSO are needed to provide the global coverage. Section 4.2 will analyze the coverage area, and the minimum number of satellites required to cover the earth surface.
2. The point-to-point communication between satellite and ground station will not be static. Mobility issue has to be considered. Detailed analysis will be done in Section 5.1.

3. The inter-satellite communication is needed, and all satellites need to form a network. details are described in Section 5.2.

In addition to above context, Section 7 will address the problem and requirements when satellite constellation is joining Internet.

As the 1st satellite constellation company in history, the SpaceX/StarLink will be inevitably mentioned in the draft. But it must be noted that all information about SpaceX/StarLink in the draft are from public. Authors of the draft have no relationship or relevant inside knowledge of SpaceX/Starlink.

4. Basics of Satellite Constellation

This section will introduce some basics for satellite such as orbit parameters, coverage estimation, minimum number of satellite and orbit plane required, real deployments.

4.1. Satellite Orbit

The orbit of a satellite can be either circular or elliptic, it can be described by following Keplerian elements [KeplerianElement]:

1. Inclination (i)
2. Longitude of the ascending node (Ω)
3. Eccentricity (e)
4. Semimajor axis (a)
5. Argument of periapsis (ω)
6. True anomaly (ν)

For a circular orbit, two parameters, Inclination and Longitude of the ascending node, will be enough to describe the orbit.

4.2. Coverage of LEO and VLEO Satellites and Minimum Number Required

The coverage of a satellite is determined by many physical factors, such as spectrum, transmitter power, the antenna size, the altitude of satellite, the air condition, the sensitivity of receiver, etc. EIRP could be used to measure the real power distribution for coverage. It is not deterministic due to too many variants in a real environment. The alternative method is to use the minimum elevation angle from user terminals or gateways to a satellite. This is easier and more deterministic. [SpaceX-Non-GEO] has suggested originally

the minimum elevation angle of 35 degrees and deduced the radius of the coverage area is about 435km and 1230km for VLEO (altitude 335.9km) and LEO (altitude 1150km) respectively. The details about how the coverage is calculated from the satellite elevation angle can be found in [Satellite-coverage].

Using this method to estimate the coverage, we can also estimate the minimum number of satellites required to cover the earth surface.

It must be noted, SpaceX has recently reduced the required minimum elevation angle from 35 degrees to 25 degrees. The following analysis still use 35 degrees.

Assume there is multiple orbit planes with the equal angular interval across the earth surface (The Longitude of the ascending node for sequential orbit plane is increasing with a same angular interval). Each orbit plane will have:

1. The same altitude.
2. The same inclination of 90 degree.
3. The same number of satellites.

With such deployment, all orbit planes will meet at north and south pole. The density of satellite is not equal. Satellite is more dense in the space above the polar area than in the space above the equator area. Below estimations are made in the worst covered area, or the area of equator where the satellite density is the minimum.

Figure 1 illustrates the coverage area on equator area, and each satellite will cover one hexagon area. The figure is based on plane geometry instead of spherical geometry for simplification, so, the orbit is parallel approximately.

Figure 2 shows how to calculate the radius (R_c) of coverage area from the satellite altitude (A_s) and the elevation angle (b).

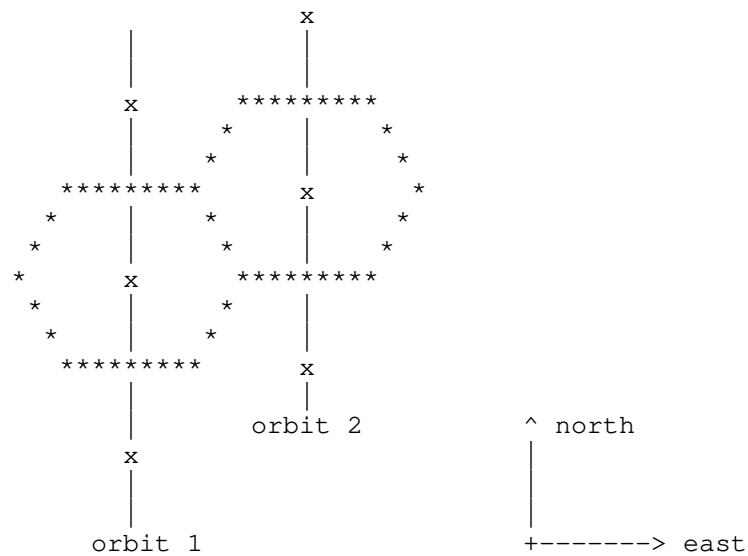


Figure 1: Satellite coverage on ground

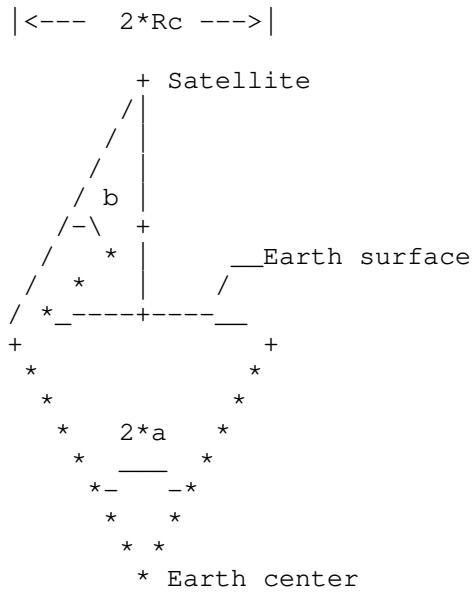


Figure 2: Satellite coverage estimation

- x The vertical projection of satelllite to Earth
- Re The radius of the Earth, Re=6378(km)

As The altitude of a satellite

Rc The radius (arc length) of the coverage, or, the arc length of hexagon center to its 6 vertices. $Rc = Re * (a * \pi) / 180$

a The cap angle for the coverage area (the RC arc). $a = \arccos((Re / (Re + As)) * \cos(b)) - b$.

b The least elevation angle that a ground station or a terminal can communicate with a satellite, $b = 35$ degree.

Ns The minimum number of satellites on one orbit plane, it is equal to the number of the satellite's vertical projection on Earth, so, $Ns = 180 / (a * \cos(30))$

No The minimum number of orbit (with same inclination), it is equal to the number of the satellite orbit's vertical projection, so, $No = 360 / (a * (1 + \sin(30)))$

For a example of two type of satelllite LEO and VEO, the coverages are calculated as in Table 1:

Parameters	VLEO1	VLEO2	LEO1	LEO2
As (km)	335.9	450	1100	1150
a (degree)	3.907	5.078	10.681	11.051
Rc (km)	435	565	1189	1230
Ns	54	41	20	19
No	62	48	23	22

Table 1: Satellite coverage estimation for LEO and VLEO examples

4.3. Real Deployment of LEO and VLEO for Satellite Network

Obviously, the above orbit parameter setup is not optimal since the sky in the polar areas will have the highest density of satellite.

In the real deployment, to provide better coverage for the areas with denser population, to get redundancy and better signal quality, and to make the satellite distance within the range of inter-satellite communication (2000km [Laser-communication-range]), more than the minimum number of satellites are launched. For example, different orbit planes with different inclination/altitude are used.

Normally, all satellites are grouped by orbit planes, each group has a number of orbit planes and each orbit plane has the same orbit parameters, so, each orbit in the same group will have:

1. The same altitude
2. The same inclination, but the inclination is less than 90 degrees. This will result in the empty coverage for polar areas and better coverage in other areas. See the orbit picture for phrase 1 for [StarLink].
3. The same number of satellites
4. The same moving direction for all satellites

The proposed deployment of SpaceX can be seen in [SpaceX-Non-GEO] for StarLink.

The China constellation deployment and orbit parameters can be seen in [China-constellation].

5. Communications for Satellite Constellation

Unlike the communication on ground, the communication for satellite constellation is much more complicated. There are two mobility aspects, one is between ground-station and satellite, another is between satellites.

In the traditional mobility communication system, only terminal is moving, the mobile core network including base station, front haul and back haul are static, thus an anchor point, i.e., PGW in 4G or UPF in 5G, can be selected for the control of mobility session. Unfortunately, when satellite constellation joins the static network system of Internet on ground, there is no such anchor point can be selected since the whole satellite constellation network is moving.

Another special aspect that can impact the communication is that the fast moving speed of satellite will cause frequent changes of communication peers and link states, this will make big challenges to the network side for the packet routing and delivery, session control and management, etc.

5.1. Dynamic Ground-station-Satellite Communication

All satellites are moving and will lead to the communication between ground station and satellite can only last a certain period of time. This will greatly impact the technologies for the satellite networking. Below illustrates the approximate speed and the time for a satellite to pass through its covered area.

In Table 2, VLEO1 and LEO3 have the lowest and highest altitude respectively, VLEO2 is for the highest altitude for VLEO. We can see that longest communication time of ground-station-satellite is less than 400 seconds, the longest communication time for VLEO ground-station-satellite is less than 140 seconds.

The "longest communication time" is for the scenario that the satellite will fly over the receiver ground station exactly above the head, or the ground station will be on the diameter line of satellite coverage circular area, see Figure 1.

Re The radius of the Earth, $Re=6378(km)$

As The altitude of a satellite

AL The arc length(in km) of two neighbor satellite on the same orbit plane, $AL=2*\cos(30)*(Re+As)*(a*pi)/180$

SD The space distance(in km) of two neighbor satellite on the same orbir plane, $SD=2*(Re+As)*\sin(AL/(2*(Re+As)))$.

V the velocity (in m/s) of satellite, $V=\sqrt{G*M/(Re+As)}$

G Gravitational constant, $G=6.674*10^{(-11)}(m^3/(kg*s^2))$

M Mass of Earth, $M=5.965*10^{24}(kg)$

T The time (in second) for a satellite to pass through its cover area, or, the time for the station-satellite communication. $T=ALs/V$

Parameters	VLEO1	VLEO2	LEO1	LEO2	LEO3
As (km)	335.9	450	1100	1150	1325
a (degree)	3.907	5.078	10.681	11.051	12.293
AL (km)	793	1048	2415	2515	2863
SD (km)	792.5	1047.2	2404	2503.2	2846.1
V (km/s)	7.7	7.636	7.296	7.272	7.189
T (s)	103	137	331	346	398

Table 2: The time for the ground-station-satellite communication

5.2. Dynamic Inter-satellite Communication

5.2.1. Inter-satellite Communication Overview

In order to form a network by satellites, there must be an inter-satellite communication. Traditionally, inter-satellite communication uses the microwave technology, but it has following disadvantages:

1. Bandwidth is limited and only up to 600M bps [Microwave-vs-Laser-communication].
2. Security is a concern since the microwave beam is relatively wide and it is easy for 3rd party to sniff or attack.
3. Big antenna size.
4. Power consumption is high.
5. High cost per bps.

Recently, laser is used for the inter-satellite communication, it has following advantages, and will be the future for inter-satellite communication.

1. Higher bandwidth and can be up to 10G bps [Microwave-vs-Laser-communication].

2. Better security since the laser beam size is much narrower than microwave, it is harder for sniffing.
3. The size of optical lens for laser is much smaller than microwave's antenna size.
4. Power saving compared with microwave.
5. Lower cost per bps.

The range for satellite-to-satellite communications has been estimated to be approximately 2,000 km currently [Laser-communication-range].

From Table 2, we can see the Space Distance (SD) for some LEO (altitude over 1100km) are exceeding the ceiling of the range of laser communication, so, the satellite and orbit density for LEO need to be higher than the estimation values in the Table 1.

Assume the laser communication is used for inter-satellite communication, then we can analyze the lifetime of inter-satellite communication when satellites are moving. The Figure 3 illustrates the movement and relative position of satellites on three orbits. The inclination of orbit planes is 90 degrees.

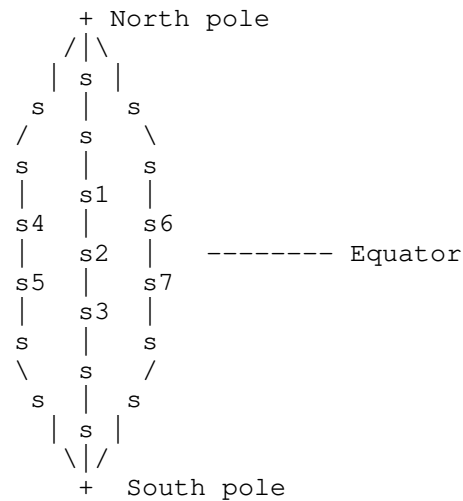
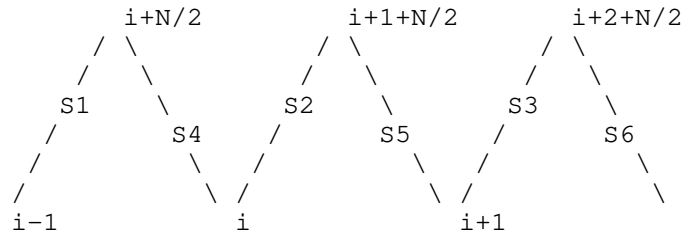


Figure 3: Satellite movement

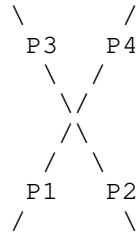
There are four scenarios:

1. For satellites within the same orbit
The satellites in the same orbit will move to the same direction with the same speed, thus the interval between satellites is relatively steady. Each satellite can communicate with its front and back neighbor satellite as long as satellite's orbit is maintained in its life cycle. For example, in Figure 3, s2 can communication with s1 and s3.
2. For satellites between neighbor orbits in the same group at non-polar areas
The orbits for the same group will share the same orbit altitude and inclination. So, the satellite speed in different orbit are also same, but the moving direction may be same or different. Figure 4 illustrates this scenario. When the moving direction is the same, it is similar to the scenario 1, the relative position of satellites in different orbit are relatively steady as long as satellite's orbit is maintained in its life cycle. When the moving direction is different, the relative position of satellites in different orbit are un-steady, this scenario will be analyzed in more details in Section 5.2.2.
3. For satellites between neighbor orbits in the same group at polar areas
For satellites between neighbor orbits with the same speed and moving direction, the relative position is steady as described in #2 above, but the steady position is only valid at areas other than polar area. When satellites meet in the polar area, the relative position will change dramatically. Figure 5 shows two satellites meet in polar area and their ISL facing will be swapped. So, if the range of laser pointing angle is 360 degrees and tracking technology supports, the ISL will not be flipping after passing polar area; Otherwise, the link will be flipping and inter-satellite communication will be interrupted.
4. For satellites between different orbits in the different group
The orbits for the different group will have different orbit altitude, inclination and speed. So, the relative position of satellite is not static. The inter-satellite communication can only last for a while when the distance between two satellite is within the limit of inter-satellite communication, that is 2000km for laser [Laser-communication-range], this scenario will be analyzed in more details in Section 5.2.3



- * The total number of orbit planes are N
- * The number ($i-1, i, i+1, \dots$) represents the Orbit index
- * The bottom numbers ($i-1, i, i+1$) are for orbit planes on which satellites (S1, S2, S3) are moving from bottom to up.
- * The top numbers ($i+N/2, i+1+N/2, i+2+N/2$) are for orbit planes on which satellites (S4, S5, S6) are moving from up to bottom.

Figure 4: Two satellites with same altitude and inclination (i) move in the same or opposite direction



- * Two satellites S1 and S2 are at position P1 and P2 at time $T1$
- * S1's right facing ISL connected to S2's left facing ISL
- * S1 and S2 move to the position P4 and P3 at time $T2$
- * S1's left facing ISL connected to S2's right facing ISL

Figure 5: Two satellites meeting in the polar area will change its facing of ISL

5.2.2. Satellites on Adjacent Orbit Planes with Same Altitude

For satellites on different orbit planes with same altitude, the estimation of the lifetime when two satellite can communicate are as follows.

Figure 6 illustrates a general case that two satellites move and intersect with an angle A .

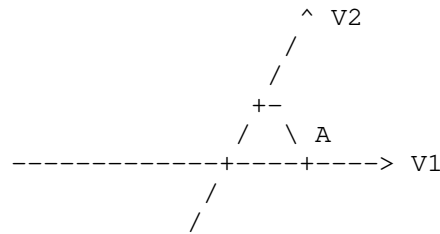


Figure 6: Two satellites (speed vector V1 and V2) intersect with angle A

More specifically, for orbit planes with the inclination angle i , Figure 7 illustrates two satellites move in the opposite direction and intersect with an angle $2*i$.

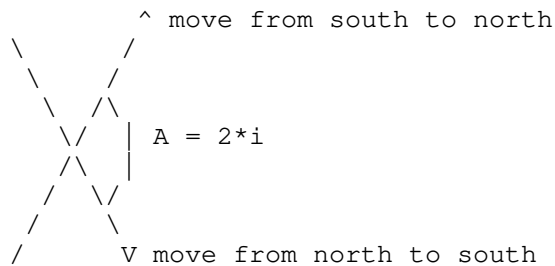


Figure 7: Two satellites with same altitude and inclination (i) intersect with angle $A=2*i$

Follows are the math to calculate the lifetime of communication. Table 3 are the results using the math for two satellites with different altitudes and different inclination angles.

D1 The laser communication limit, $D1=2000\text{km}$
[Laser-communication-range]

A The angle between two orbit's vertical projection on Earth.
 $A=2*i$

V1 The speed vector of satellite on orbit1

V2 The speed vector of satellite on orbit2

$|V|$ the magnitude of the difference of two speed vector V1 and V2, $|V|=|V1-V2|=\sqrt{(V1-V2*\cos(A))^2+(V2*\sin(A))^2}$. For satellites with the same altitude and inclination angle i , $V1=V2$, so, $|V|=V1*\sqrt{2-2*\cos(2*i)}=2V1*\sin(i)$

T The lifetime two satellites can communicate, or the time of two satellites' distance is within the range of communication, $T = 2 \cdot D_l / |V|$.

i (degree)	80	80	65	65	50	50
Alt (km)	500	800	500	800	500	800
V (km/s)	14.98	14.67	13.79	13.5	11.66	11.41
T(s)	267	273	290	296	343	350

Table 3: The lifetime of communication for two LEOs (with two altitudes and three inclination angles)

5.2.3. Satellites on Adjacent Orbit Planes with Different Altitude

For satellites on different orbit planes with different altitude, the estimation of the lifetime when two satellite can communicate are as follows.

Figure 8 illustrates two satellites (with the altitude difference D_a) move and intersect with an angle A .

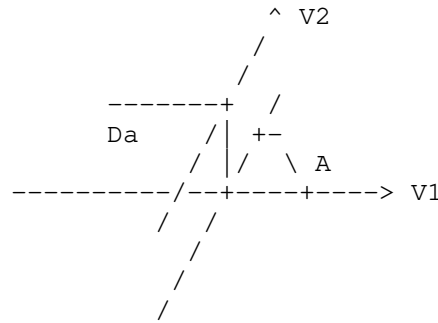


Figure 8: Satellite (speed vector V_1 and V_2 , Altitude difference D_a) intersects with Angle A

Follows are the math to calculate the lifetime of communication

D_l The laser communication limit, $D_l=2000\text{km}$
[Laser-communication-range]

D_a Altitude difference (in km) for two orbit planes

A The angle between two orbit's vertical projection on Earth

V1 The speed vector of satellite on orbit 1

V2 The speed vector of satellite on orbit 2

$|V|$ the magnitude of the difference of two speed vector V1 and V2, $|v| = |V1 - V2| = \sqrt{(V1 - V2 \cos(A))^2 + (V2 \sin(A))^2}$

T The lifetime two satellites can communicate, or the time of two satellites' distance is within the range of communication, $T = 2 \times \sqrt{D1^2 - Da^2} / |V|$

Using formulas above, below is the estimation for the life of communication of two satellites when they intersect. Table 4 and Table 5 are for two VLEOs with the difference of 114.1km for altitude. (VLEO1 and VLEO2 on Table 2). Table 6 and Table 7 are for two LEOs with the difference of 175km for altitude (LEO2 and LEO3 on Table 2).

Parameters	VLEO1	VLEO2
As (km)	335.9	450
V (km/s)	7.7	7.636

Table 4: Two VLEO with different altitude and speed

A (degree)	0	10	45	90	135	180
$ V $ (km/s)	0.065	1.338	5.869	10.844	14.169	15.336
T(s)	61810	2984	680	368	282	260

Table 5: Two VLEO intersects with different angle and the life of communication

Parameters	LEO1	LEO2
As (km)	1150	1325
V (km/s)	7.272	7.189

Table 6: Two LEO with different altitude and speed

A (degree)	0	10	45	90	135	180
V (km/s)	0.083	1.263	5.535	10.226	13.360	14.461
T(s)	47961	3155	720	390	298	276

Table 7: Two LEO intersects with different angle and the life of communication

6. Use Satellite Network for Internet

Since there is no complete satellite network established yet, all following analysis is based on the predictions from the traditional GEO communication. The analysis also learnt how other type of network has been used in Internet, such as Broadband access network, Mobile access network, Enterprise network and Service Provider network.

As a criteria to be part of Internet, any device connected to any satellite should be able to communicate with any public IP4 or IPv6 address in Internet. There could be three types of methods to deliver IP packet from source to destination by satellite:

1. Data packet is relayed between ground station and satellite.
For this method, there is no inter-satellite communication and networking. Data packet is bounced once or couple times between ground stations and satellites until the packet arrives at the destination in Internet.
2. Data packet is delivered by inter-satellite networking.
For this method, the data packet traverses with multiple satellites and inter-satellite networking is used to deliver the packet to the destination in Internet.

3. Both satellite relay and inter-satellite networking are used. For this method, the data packet is relayed in some segments and traverse with multiple satellites in other segments. It is a combination of the method 1 and method 2.

Using the above methods, follows are typical deployment scenarios that a Satellite network is integrated with Internet:

1. The end user terminal access Internet through satellite relay (Figure 9 for one satellite relay, Figure 10 for multiple satellite relay).
2. The end user terminal access Internet through inter-satellite-networking (Figure 11).
3. The local network access Internet through satellite relay (Figure 12 for one satellite relay, Figure 13 for multiple satellite relay).
4. The local network access Internet through inter-satellite-networking (Figure 14).
5. The End user terminal or local network access Internet through satellite network and Mobile Access Network, From mobile access network to satellite network or From satellite network to mobile access network, Satellite network includes inter satellite network and relay network (Figure 15 for mobile access network to satellite network, Figure 16 for satellite netowk to mobile access network).

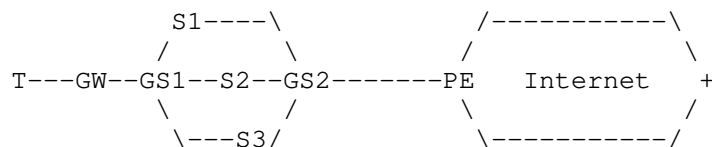


Figure 9: End user terminal access Internet through one satellite relay

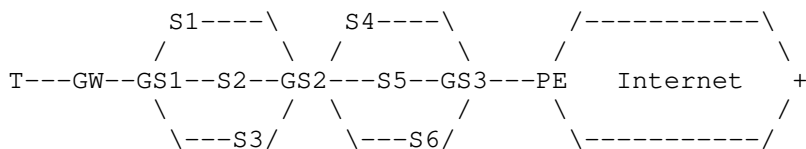


Figure 10: End user terminal access Internet through multiple satellite relay

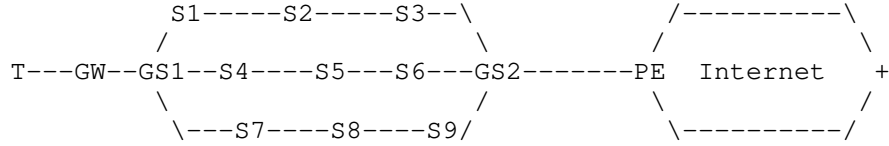


Figure 11: End user terminal access Internet through inter-satellite-networking

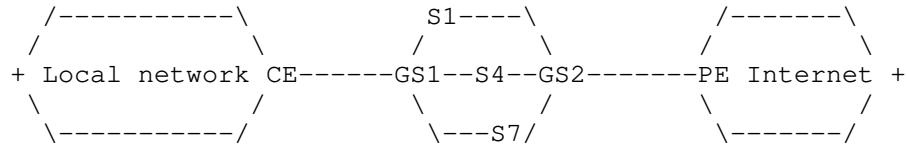


Figure 12: Local network access Internet through one satellite relay

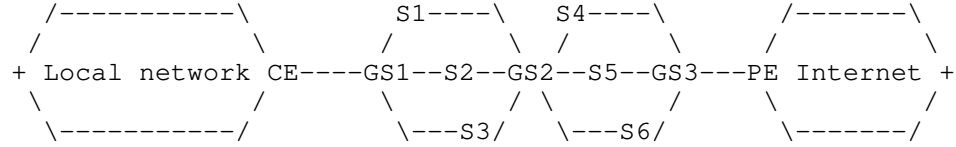


Figure 13: Local network access Internet through multiple satellite relay

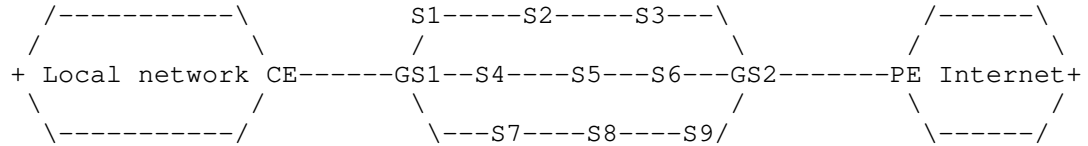


Figure 14: Local network access Internet through inter-satellite-networking

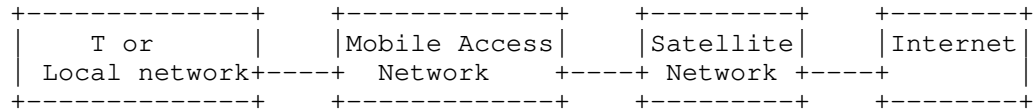


Figure 15: End user terminal or local network access Internet through Mobile Access Network and Satellite Network

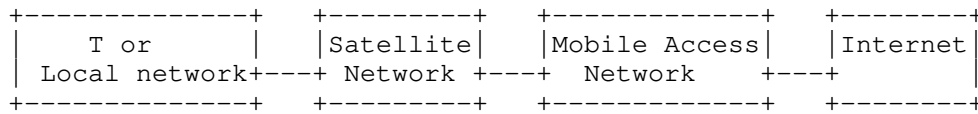


Figure 16: End user terminal or local network access Internet through Satellite Network and Mobile Access Network

In above Figure 9 to Figure 16, the meaning of symbols are as follows:

T	The end user terminal
GW	Gateway router
GS1, GS2, GS3	Ground station with L2/L3 routing/switch functionality.
S1 to S9	Satellites
PE	Provider Edge Router
CE	Customer Edge Router

7. Problems and Requirements for Satellite Constellation for Internet

As described in Section 6, satellites in a satellite constellation can either relay internet traffic or multiple satellites can form a network to deliver internet traffic. More detailed analysis are in following sub sections. There might have multiple solutions for each method described in Section 6, following contexts only discuss the most plausible solution from networking perspectives.

Section 7.1 will list the common problems and requirements for both satellite relay and satellite networking.

Section 7.2 and Section 7.3 will describe key problems, requirement and potential solution from the networking perspective for these two cases respectively.

7.1. Common Problems and Requirements

For both satellite relay and satellite networking, satellite-ground-station must be used, so, the problems and requirements for the satellite-ground-station communication is common and will apply for both methods.

When one satellite is communicating with ground station, the satellite only needs to receive data from uplink of one ground station, process it and then send to the downlink of another ground station. Figure 9 illustrates this case. Normally microwave is used for both links.

Additionally, from the coverage analysis in Section 4.2 and real deployment in Section 4.3, we can see one ground station may communicate with multiple satellites. Similarly, one satellite may communicate with multiple ground stations. The characters for satellite-ground-station communication are:

1. Satellite-ground-station communication is P2MP.
Since microwave physically is the carrier of broadcast communication, one satellite can send data while multiple ground stations can receive it. Similarly, one ground station can send data and multiple satellites can receive it.
2. Satellite-ground-station communication is in open space and not secure.
Since electromagnetic fields for microwave physically are propagating in open space. The satellite-ground-station communication is also in open space. It is not secure naturally.
3. Satellite-ground-station communication is not steady.
Since the satellite is moving with high speed, from Section 5.1, the satellite-ground-station communication can only last a certain period of time. The communication peers will keep changing.
4. Satellite-to-Satellite communication is not steady.
For some satellites, even they are in the same altitude and move in the same speed, but they move in the opposite direction, from Section 5.2.2, the satellite-to-satellite communication can only last a certain period of time. The communication peers will keep changing.
5. Satellite-to-Satellite distance is not steady.
For satellites with the same altitude and same moving direction, even their relative position is steady, but the distance between satellites are not steady. This will lead to the inter-satellite-communication's bandwidth and latency keep changing.

6. Satellite physical resource is limited.
Due to the weight, complexity and cost constraint, the physical resource on a satellite, such as power supply, memory, link speed, are limited. It cannot be compared with the similar device on ground. The design and technology used should consider these factors and take the appropriate approach if possible.

The requirements of satellite-ground-station communication are:

- R1. The bi-directional communication capability
Both satellites and ground stations have the bi-directional communication capability
- R2. The identifier for satellites and ground stations
Satellites and ground stations should have Ethernet and/or IP address configured for the device and each link. More detailed address configuration can be seen in each solution.
- R3. The capability to decide where the IP packet is forwarded to.
In order to send Internet traffic or IP data to destination correctly, satellites and ground station must have Ethernet hub or switching or IP routing capability. More detailed capability can be seen in each solution.
- R4. The protocol to establish the satellite-ground-station communication.
For security and management purpose, the satellite-ground-station communication is only allowed after both sides agree through a protocol. The protocol should be able to establish a secured channel for the communication when a new communication peer comes up. Each ground station should be able to establish multiple channels to communicate with multiple satellites. Similarly, each satellite should be able to establish multiple channels to communicate to multiple ground stations.
- R5. The protocol to discover the state of communication peer.
The discover protocol is needed to detect the state of communication peer such as peer's identity, the state of the peer and other info of the peer. The protocol must be running securely without leaking the discovered info.
- R6. The internet data packet is forwarded securely.
When satellite or ground station is sending the IP packet to its peer, the packet must be relayed securely without leaking the user data.

R7. The internet data packet is processed efficiently on satellite

Due to the resource constraint on a satellite, the packet may need more efficient mechanism to be processed on satellite. The process on satellite should be very minimal and offloaded to ground as much as possible.

7.2. Satellite Relay

One of the reasons to use satellite constellation for internet access is it can provide shorter latency than using the fiber underground. But using ISL for inter-satellite communication is the premise for such benefit in latency. Since the ISL is still not mature and adopted commercially, satellite relay is a only choice currently for satellite constellation used for internet access. In [UCL-Mark-Handley], detailed simulations have demonstrated better latency than fiber network by satellite relay even the ISL is not present.

7.2.1. One Satellite Relay

One satellite relay is the simplest method for satellite constellation to provide Internet service. By this method, IP traffic will be relayed by one satellite to reach the DGS and go to Internet.

The solution option and associated requirements are:

S1. The satellite only does L1 relay or the physical signal process.

For this solution, a satellite only receives physical signal, amplify it and broadcast to ground stations. It has no further process for packet, such as L2 packet compositing and processing, etc. All packet level work is done only at ground station. The requirements for the solution are:

R1-1. SGS and BGS are configured as IP routing node. Routing protocol is running in SGS and BGS

SGS and BGS is a IP peer for a routing protocol (IGP or BGP). SGS will send internet traffic to DGS as next hop through satellite uplink and downlink.

R1-2. DGS must be connected with Internet.

DGS can process received packet from satellite and forward the packet to the destination in Internet.

In addition to the above requirements, following problem should be solved:

P1-1. IP continuity between two ground stations

This problem is that two ground stations are connected by one satellite relay. Since the satellite is moving, the IP continuity between ground stations is interrupted by satellite changing periodically. Even though this is not killing problem from the view point that IP service traditionally is only a best effort service, it will benefit the service if the problem can be solved. Different approaches may exist, such as using hands off protocols, multipath solutions, etc.

S2. The satellite does the L2 relay or L2 packet process.

For this solution, IP packet is passing through individual satellite as an L2 capable device. Unlike in the solution S1, satellite knows which ground station it should send based on packet's destination MAC address after L2 processing. The advantage of this solution over S1 is it can use narrower beam to communicate with DGS and get higher bandwidth and better security. The requirements for the solution are:

R2-1. Satellite must have L2 bridge or switch capability

In order to forward packet to properly, satellite should run some L2 process such as MAC learning, MAC switching. The protocol running on satellite must consider the fast movement of satellite and its impact to protocol convergence, timer configuration, table refreshment, etc.

R2-2. same as R1-1 in S1

R2-3. same as R1-2 in S1

In addition to the above requirements, the problem P1-1 for S1 should also apply.

7.2.2. Multiple Satellite Relay

For this method, packet from SGS will be relayed through multiple intermediate satellites and ground station until reaching a DGS.

This is more complicated than one satellite relay described in Section 7.2.1.

One general solution is to configure both satellites and ground-stations as IP routing nodes, proper routing protocols are running in this network. The routing protocol will dynamically determine forwarding path. The obvious challenge for this solution is that all links between satellite and ground station are not static, according to the analysis in Section 5.1, the lifetime of each link may last

only couple of minutes. This will result in very quick and constant topology changes in both link state and IP adjacency, it will cause the distributed routing algorithms may never converge. So this solution is not feasible.

Another plausible solution is to specify path statically. The path is composed of a serials of intermediate ground stations plus SGS and DGS. This idea will make ground stations static and leave the satellites dynamic. It will reduce the fluctuation of network path, thus provide more steady service. One variant for the solution is whether the intermediate ground stations are connected to Internet. Separated discussion is as below:

S1. Manual configuring routing path and table

For this solution, the intermediate ground stations and DGS are specified and configured manually during the stage of network planning and provisioning. Following requirements apply:

R1-1. Specify a path from SGS to DGS via a list of intermediate ground stations.

The specified DGS must be connected with internet. Other specified intermediate ground stations does not have to

R1-2. All Ground stations are configured as IP routing node. Static routing table on all ground stations must be pre-configured, the next hop of routes to Internet destination in any ground station is configured to going through uplink of satellite to the next ground station until reaching the DGS.

R1-3. All Satellites are configured as either L1 relay or L2 relay.

The Satellite can be configured as L1 relay or L2 relay described in S1 and S2 respectively in Section 7.2.1

In addition to the above requirements, the problem P1-1 in Section 7.2.1 should also apply.

S2. Automatic decision by routing protocol.

This solution is only feasible after the IP continuity problem (P1-1 in Section 7.2.1) is solved. Following requirements apply:

R2-1. All Ground stations are configured as IP routing node.
Proper routing protocols are configured as well.

The satellite link cost is configured to be lower than the ground link. In such a way, the next hop of routes for the IP forwarding to Internet destination in any ground station will be always going through the uplink of satellite to the next ground station until reaching the DGS.

R2-2. All Satellites are configured as either L1 relay or L2 relay.

The Satellite can be configured as L1 relay or L2 relay described in S1 and S2 respectively in Section 7.2.1

In addition to the above requirements, the problem P1-1 in Section 7.2.1 should also apply.

7.3. Satellite Networking

In the draft, satellite Network is defined as a network that satellites are inter-connected by inter-satellite links (ISL). One of the major difference of satellite network with the other type of network on ground (telephone, fiber, etc.) is its topology and links are not stationary, some new issues have to be considered and solved. Follows are the factors that impact the satellite networking.

7.3.1. L2 or L3 network

The 1st question to answer is should the satellite network be configured as L2 or L3 network? As analyzed in Section 4.2 and Section 4.3, since there are couple of hundred or over ten thousand satellites in a network, L2 network is not a good choice, instead, L3 or IP network is more appropriate for such scale of network.

7.3.2. Inter-satellite-Link Lifetime

If we assume the orbit is circular and ignore other trivial factors, the satellite speed is approximately determined by the orbit altitude as described in the Section 5.1. The satellite orbit can determine if the dynamic position of two satellites is within the range of the inter-satellite communication. That is 2000km for laser communication [Laser-communication-range] by Inter Satellite Laser Link (ISLL).

When two satellites' orbit planes belong to the same group, or two orbit planes share the same altitude and inclination, and when the satellites move in the same direction, the relative positions of two satellites are relatively stationary, and the inter-satellite communication is steady. But when the satellites move in the

opposite direction, the relative positions of two satellites are not stationary, the communication lifetime is couple of minutes. The Section 5.2.2 has analyzed the scenario.

When two satellites' orbit planes belong to the different group, or two orbit planes have different altitude, the relative position of two satellite are unstable, and the inter-satellite communication is not steady. As described in Section 5.2, The life of communication for two satellites depends on the following parameters of two satellites:

1. The speed vectors.
2. The altitude difference
3. The intersection angle

From the examples shown in Table 4 to Table 7, we can see that the lifetime of inter-satellite communication for the different group of orbit planes are from couple of hundred seconds to about 18 hours. This fact will impact the routing technologies used for satellite network and will be discussed in Section 7.3.3.

7.3.3. Problems for Traditional Routing Technologies

When the satellite network is integrated with Internet by traditional routing technologies, following provisioning and configuration (see Figure 17) will apply:

1. The ground stations connected to local network and internet are treated as PE router for satellite network (called PE_GS1 and PE_GS2 in the following context), and all satellites are treated as P router.
2. All satellites in the network and ground stations are configured to run IGP.
3. The eBGP is configured between PE_GS and its peered network's PE or CE.

The work on PE_GS1 are:

- * The local network routes are received at PE_GS1 from CE by eBGP. The routes are redistributed to IGP and then IGP flood them to all satellites. (Other more efficient methods, such as iBGP or BGP reflectors are hard to be used, since the satellite is moving and there is no easy way to configure a full meshed iBGP session for all satellites, or configure one satellite as BGP reflector in satellite network.)
- * The internet routes are redistributed from IGP to eBGP running on PE_GS1, and eBGP will advertise them to CE.

The work on PE_GS2 are:

- * The Internet routes are received at PE_GS2 from PE by eBGP. The routes are redistributed to IGP and then IGP flood them to all satellites. (Similar as in PE_GS1, Other more efficient methods, such as iBGP or BGP reflector cannot be used.)
- * The local network routes are redistributed from IGP to eBGP running on PE_GS2, and eBGP will advertise them to Internet.

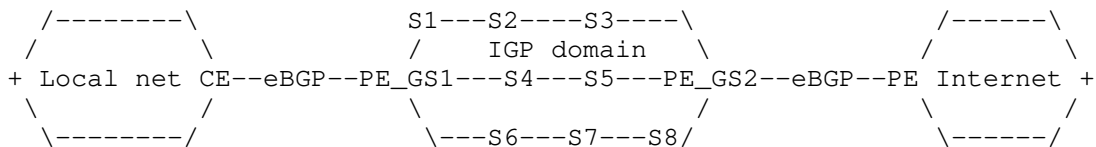


Figure 17: Local access Internet through inter-satellite-networking

Local access Internet through inter-satellite-networking

On PE-GS1, due to the fact that IGP link between PE_GS1 and satellite is not steady; this will lead to following routing activity:

1. When one satellite is connecting with PE_GS1, the satellite and PE_GS1 form a IGP adjacency. IGP starts to exchange the link state update.
2. The local network routes received by eBGP in PE_GS1 from CE are redistributed to IGP, and IGP starts to flood link state update to all satellites.
3. Meanwhile, the Internet routes learnt from IGP in PE_GS1 will be redistributed to eBGP. eBGP starts to advertise to CE.
4. Every satellite will update its routing table (RIB) and forwarding table (FIB) after IGP finishes the SPF algorithm.

5. When the satellite is disconnecting with PE-GS1, the IGP adjacency between satellite and PE_GS1 is gone. IGP starts to exchange the link state update.
6. The routes of local network and satellite network that were redistributed to IGP in step 2 will be withdrawn, and IGP starts to flood link state update to all satellites.
7. Meanwhile, the Internet routes previously redistributed to eBGP in step 3 will also be withdrawn. eBGP starts to advertise route withdraw to CE.
8. Every satellite will update its routing table (RIB) and forwarding table (FIB) after the SPF algorithm.

Similarly on PE_GS2, due to the fact that IGP link between PE_GS2 and satellite is not steady; this will lead to following routing activity:

1. When one satellite is connecting with PE_GS2, the satellite and PE_GS2 form a IGP adjacency. IGP starts to exchange the link state update.
2. The Internet routes previously received by eBGP in PE_GS2 from PE are redistributed to IGP, IGP starts to flood the new link state update to all satellites.
3. Meanwhile, the routes of local network and satellite network learnt from IGP in PE_GS2 will be redistributed to eBGP. eBGP starts to advertise to Internet peer PE.
4. Every satellite will update its routing table (RIB) and forwarding table (FIB) after IGP finishes the SPF algorithm.
5. When the satellite is disconnecting with PE-GS2, the IGP adjacency between satellite and PE_GS2 is gone. IGP starts to exchange the link state update.
6. The internet routes previously redistributed to IGP in step 2 will be withdrawn, and IGP starts to flood link state update to all satellites
7. Meanwhile, the routes of local network and satellite network previously redistributed to eBGP in step 3 will also be withdrawn. eBGP starts to advertise route withdraw to PE.
8. Every satellite will update its routing table (RIB) and forwarding table (FIB) after the SPF algorithm.

For the analysis of detailed events above, the estimated time interval between event 1 and 5 for PE_GS1 and PE_GS2 can use the analysis in Section 5.1. For example, it is about 398s for LEO and 103s for VLEO. Within this time interval, the satellite network including all satellites and two ground stations must finish the works from 1 to 4 for PE_GS1 and PE_GS2. The normal internet IPv6 and IPv4 BGP routes size are about 850k v4 routes + 100K v6 routes [BGP-Table-Size]. There are couple critical problems associated with the events:

P1. Frequent IGP update for its link cost

Even for satellites in different orbit with the steady relative positions, the distance between satellites is keep changing. If the distance is used as the link cost, it means the IGP has to update the link cost frequently. This will make IGP keep running and update its routing table.

P2. Frequent IGP flooding for the internet routes

Whenever the IGP adjacency changes (step 1 and 5 for PE_GS2), it will trigger the massive IGP flooding for the link state update for massive internet routes learnt from eBGP. This will result in the IGP re-convergency, RIB and FIB update.

P3. Frequent BGP advertisement for the internet routes

Whenever the IGP adjacency changes (step 3 and 7 for PE_GS1), it will trigger the massive BGP advertisement for the internet routes learnt from IGP. This will result in the BGP re-convergency, RIB and FIB update. BGP convergency time is longer than IGP. The document [BGP-Converge-Time1] has shown that the BGP convergence time varies from 50sec to couple of hundred seconds. The analysis [BGP-Converge-Time2] indicated that per entry update takes about 150us, and it takes $O(75s)$ for 500k routes, or $O(150s)$ for 1M routes.

P4. More frequent IGP flooding and BGP update in whole satellite network

To provide the global coverage, a satellite constellation will have many ground stations deployed. For example, StarLink has applied for the license for up to one million ground stations [StarLink-Ground-Station-Fcc], in which, more than 50 gateway ground stations (equivalent to the PE_GS2) have been registered [SpaceX-Ground-Station-Fcc] and deployed in U.S. [StarLink-GW-GS-map]. It is expected that the gateway ground station will grow quickly to couple of thousands [Tech-Comparison-LEOs]. This means almost each satellite in the satellite network would have a ground station connected. , Due to the fact that all satellites are moving, many IGP adjacency changes may occur in a shorter period of time described in Section 5.1 and result in the problem P1 and P2 constantly occur.

P5. Service is not steady

Due to the problems P1 to P3, the service provider of satellite constellation is hard to provide a steady service for broadband service by using inter-satellite network and traditional routing technologies.

As a summary, the traditional routing technology is problematic for large scale inter-satellite networking for Internet. Enhancements on traditional technologies, or new technologies are expected to solve the specific issues associated with satellite networking.

8. IANA Considerations

This memo includes no request to IANA.

9. Contributors

10. Acknowledgements

11. References

11.1. Normative References

- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <<https://www.rfc-editor.org/info/rfc2328>>.

- [RFC7142] Shand, M. and L. Ginsberg, "Reclassification of RFC 1142 to Historic", RFC 7142, DOI 10.17487/RFC7142, February 2014, <<https://www.rfc-editor.org/info/rfc7142>>.
- [RFC2453] Malkin, G., "RIP Version 2", STD 56, RFC 2453, DOI 10.17487/RFC2453, November 1998, <<https://www.rfc-editor.org/info/rfc2453>>.
- [RFC7868] Savage, D., Ng, J., Moore, S., Slice, D., Paluch, P., and R. White, "Cisco's Enhanced Interior Gateway Routing Protocol (EIGRP)", RFC 7868, DOI 10.17487/RFC7868, May 2016, <<https://www.rfc-editor.org/info/rfc7868>>.

11.2. Informative References

- [KeplerianElement] "Keplerian elements", <https://en.wikipedia.org/wiki/Orbital_elements>.
- [GEO-Coverage] "Coverage of a geostationary satellite at Earth", <<https://www.planetary.org/space-images/coverage-of-a-geostationary>>.
- [Nttdocomo-6G] "NTTDP COM 6G White Paper", <https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_20200124.pdf>.
- [ITU-6G] "ITU 6G vision", <https://www.itu.int/dms_pub/itu-s/opb/itu_jnl/S-ITUJNL-JFETF.V1I1-2020-P09-PDF-E.pdf>.
- [Surrey-6G] "Surrey 6G vision", <<https://www.surrey.ac.uk/sites/default/files/2020-11/6g-wireless-a-new-strategic-vision-paper.pdf>>.
- [OSI-Model] "OSI Model", <https://en.wikipedia.org/wiki/OSI_model>.
- [StarLink] "Star Link", <<https://en.wikipedia.org/wiki/Starlink>>.
- [China-constellation] "China Constellation", <<https://www.itu.int/ITU-R/space/asreceived/Publication/DisplayPublication/23706>>.

[ESA-HydRON]

"HydRON: Fiber in the sky",
<https://www.esa.int/ESA_Multimedia/Videos/2021/04/HydRON_Fibre_in_the_sky>.

[SpaceX-Non-GEO]

"FCC report: SPACEX V-BAND NON-GEOSTATIONARY SATELLITE SYSTEM", <<https://fcc.report/IBFS/SAT-LOA-20170301-00027/1190019.pdf>>.

[Satellite-coverage]

Alan R.Washburn, Department of Operations Research, Naval Postgraduate School, "Earth Coverage by Satellites in Circular Orbit",
<<https://faculty.nps.edu/awashburn/Files/Notes/EARTHCOV.pdf>>.

[Microwave-vs-Laser-communication]

International Journal for Research in Applied Science and Engineering Technology (IJRASET), "Comparison of Microwave and Optical Wireless Inter-Satellite Links",
<<https://www.ijraset.com/files/serve.php?FID=7815>>.

[Laser-communication-range]

"Interferometric optical communications can potentially lead to robust, secure, and naturally encrypted long-distance laser communications in space by taking advantage of the underlying physics of quantum entanglement.",
<<https://www.laserfocusworld.com/optics/article/16551652/interferometry-quantum-entanglement-physics-secures-spacetospace-interferometric-communications>>.

[BGP-Table-Size]

"BGP in 2020 - BGP table",
<<https://blog.apnic.net/2021/01/05/bgp-in-2020-the-bgp-table/>>.

[BGP-Converge-Time1]

"BGP in 2020 - BGP Update Churn",
<<https://labs.apnic.net/?p=1397>>.

[BGP-Converge-Time2]

"Bringing SDN to the Internet, one exchange point at the time",
<<https://www.cs.princeton.edu/courses/archive/fall14/cos561/docs/SDX.pdf>>.

- [StarLink-Ground-Station-Fcc]
"APPLICATION FOR BLANKET LICENSED EARTH STATIONS",
<<https://fcc.report/IBFS/SES-LIC-INTR2019-00217/1616678>>.
- [SpaceX-Ground-Station-Fcc]
"List of SpaceX applications for ground stations",
<<https://fcc.report/IBFS/Company/Space-Exploration-Technologies-Corp-SpaceX>>.
- [Tech-Comparison-LEOs]
"A Technical Comparison of Three Low Earth Orbit Satellite Constellation Systems to Provide Global Broadband",
<<http://www.mit.edu/~portillo/files/Comparison-LEO-IAC-2018-slides.pdf>>.
- [StarLink-GW-GS-map]
"StarLink gateway ground station map",
<https://www.google.com/maps/d/u/0/viewer?mid=1Hlx8jZs8vfjy60TvKgpbYs_grargieVw>.
- [UCL-Mark-Handley]
"Using ground relays for low-latency wide-area routing in megaconstellations",
<<https://discovery.ucl.ac.uk/id/eprint/10090242/1/hotnets-ucl.pdf>>.

Appendix A. Change Log

- * Initial version, 07/03/2021
- * 01 version, 10/20/2021

Authors' Addresses

Lin Han (editor)
Futurewei Technologies, Inc.
2330 Central Expy
Santa Clara, CA 95050,
United States of America

Email: lhhan@futurewei.com

Richard Li
Futurewei Technologies, Inc.
2330 Central Expy
Santa Clara, CA 95050,
United States of America

Email: rli@futurewei.com

Alvaro Retana
Futurewei Technologies, Inc.
2330 Central Expy
Santa Clara, CA 95050,
United States of America

Email: alvaro.retana@futurewei.com

Meiling Chen
China Mobile
32, Xuanwumen West
BeiJing 100053
China

Email: chenmeiling@chinamobile.com

Li Su
China Mobile
32, Xuanwumen West
BeiJing 100053
China

Email: suli@chinamobile.com

Ning Wang
University of Surrey
Guildford
Surrey, GU2 7XH
United Kingdom

Email: n.wang@surrey.ac.uk

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 7 September 2022

L. Han, Ed.
R. Li
A. Retana
Futurewei Technologies, Inc.
M. Chen
China Mobile
N. Wang
University of Surrey
6 March 2022

Satellite Semantic Addressing for Satellite Constellation
draft-lhan-satellite-semantic-addressing-01

Abstract

This document presents a semantic addressing method for satellites in satellite constellation connecting with Internet. The satellite semantic address can indicate the relative position of satellites in a constellation. The address can be used with traditional IP address or MAC address or used independently for IP routing and switching.

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 7 September 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	2
2. Terminology	3
3. Overview	4
4. Basics of Satellite Constellation and Satellite Orbit	5
4.1. Satellite Orbit	5
4.2. Satellite Constellation Compositions	6
4.3. Communication between Satellites by ISL	7
5. Addressing of Satellite	9
5.1. Indexes of Satellite	9
5.2. The Range of Satellite Indexes	12
5.3. Other Info for satellite addressing	13
5.4. Encoding of Satellite Semantic Address	14
5.5. Link Identification by Satellite Semantic Address	16
6. IANA Considerations	18
7. Contributors	18
8. Acknowledgements	18
9. References	18
9.1. Normative References	18
9.2. Informative References	19
Appendix A. Change Log	21
Authors' Addresses	21

1. Introduction

Satellite constellation technologies for Internet are emerging and expected to provide Internet service like the traditional wired network on the ground. A typical satellite constellation will have couple of thousands or over ten thousand of LEO and/or VLEO. Satellites in a constellation will be connected to adjacent satellites by Inter-Satellite-Links (ISL), and/or connected to ground station by microwave or laser links. ISL is still in research stage and will be deployed soon. This memo is for the satellite networking with the use of ISL.

The memo proposes to use some indexes to represent a satellite's orbit information. The indexes can form satellite semantic address, the address can then be embedded into IPv6 address or MAC address for IP routing and switching. The address can also be used independently if the shorter than 128-bit length of IP address is accepted. As an internal address for satellite network, it only applies to satellites that will form a constellation to transport Internet traffic between ground stations and will not be populated to Internet by BGP.

2. Terminology

LEO	Low Earth Orbit with the altitude from 180 km to 2000 km.
VLEO	Very Low Earth Orbit with the altitude below 450 km
GEO	Geosynchronous orbit with the altitude 35786 km
ISL	Inter Satellite Link
ISLL	Inter Satellite Laser Link
3D	Three Dimensional
GS	Ground Station, a device on ground connecting the satellite. In the document, GS will hypothetically provide L2 and/or L3 functionality in addition to process/send/receive radio wave. It might be different as the reality that the device to process/send/receive radio wave and the device to provide L2 and/or L3 functionality could be separated.
SGS	Source ground station. For a specified flow, a ground station that will send data to a satellite through its uplink.
DGS	Destination ground station. For a specified flow, a ground station that is connected to a local network or Internet, it will receive data from a satellite through its downlink and then forward to a local network or Internet.
L1	Layer 1, or Physical Layer in OSI model [OSI-Model]
L2	Layer 2, or Data Link Layer in OSI model [OSI-Model]

L3	Layer 3, or Network Layer in OSI model [OSI-Model], it is also called IP layer in TCP/IP model
BGP	Border Gateway Protocol [RFC4271]
IGP	Interior gateway protocol, examples of IGP include Open Shortest Path First (OSPF [RFC2328]), Routing Information Protocol (RIP [RFC2453]), Intermediate System to Intermediate System (IS-IS [RFC7142]) and Enhanced Interior Gateway Routing Protocol (EIGRP [RFC7868]).

3. Overview

For IP based satellite networking, the topology is very dynamic and the traditional IGP and BGP based routing technologies will face challenges according to the analysis in [I-D.lhan-problems-requirements-satellite-net]. From the paper, we can easily categorize satellite links as two types, steady and un-steady. For un-steady links, the link status will be flipping every couple of minutes.

Section 5.5 has more details about how to identify different links.

Some researches have been done to handle such fast changed topologies. one method to overcome the difficulties for routing with un-steady links is to only use the steady links, and get rid of un-steady links unless it is necessary. For example, for real deployment, only links between satellite and ground stations are mandatory to use, other un-steady links can be avoided in routing and switching algorithms. [Routing-for-LEO] proposed to calculate the shortest path by avoiding un-steady links in polar area and links crossing Seam line since satellites will move in the opposite direction crossing the Seam line.

Traditionally, to establish an IP network for satellites, each satellite and its interface between satellites and to ground stations have to be assigned IP addresses (IPv4 or IPv6). The IP address can be either private or public. IP address itself does not mean anything except routing prefix and interface identifier [RFC8200].

To utilize the satellite relative position for routing, it is desired that there is an easy way to identify the relative positions of different satellites and identify un-steady links quickly. The traditional IP address cannot provide such functionality unless we have the real-time processing for 3D coordinates of satellites to figure out the relative positions of each satellite, and some math calculation and dynamic database are also needed in routing algorithm

to check if a link is steady or not. This will introduce extra data exchanged for routing protocols and burden for the computation in every satellite. Considering the ISL link speed (up to 10G for 2000km) and hardware cost (Radiation-hardened semiconductor components are needed) in satellite are more constraint than for network device on ground, it is expected to simplify the routing algorithm, reduce the requirement of ISL, onboard CPU and memory.

The document proposes to form a semantic address by satellite orbit information, and then embedded it into a proper IP address. The IP address of IGP neighbors can directly tell the relative position of different satellites and if links between two satellites are steady or not.

The document does not describe the details how the semantic address is used to improve routing and switching or new routing protocols, those will be addressed in different documents.

4. Basics of Satellite Constellation and Satellite Orbit

This section will introduce some basics for satellite such as orbit parameters.

4.1. Satellite Orbit

The orbit of a satellite can be either circular or elliptic, it can be described by following Keplerian elements [KeplerianElement]:

1. Inclination (i)
2. Longitude of the ascending node (Ω)
3. Eccentricity (e)
4. Semimajor axis (a)
5. Argument of periapsis (ω)
6. True anomaly (ν)

The circular orbit is widely used by proposals of satellite constellation from different companies and countries.

For a circular orbit, we will have:

- * Eccentricity $e = 0$
- * Semimajor axis $a = \text{Altitude of satellite}$

* Argument of periapsis $\omega = 90$ degree

So, three parameters, Altitude, Inclination and Longitude of the ascending node, will be enough to describe the orbit. The satellite will move in a constant speed and True anomaly (ν) can be easily calculated after the epoch time is defined.

4.2. Satellite Constellation Compositions

One satellite constellation may be composed of many satellites (LEO and VLEO), but normally all satellites are grouped in a certain order that is never changed during the life of satellite constellation. Each satellite constellation's orbits parameters described in Section 4.1 must be approved by regulator and cannot be changed either. Follows are characters of one satellite constellation:

1. One Satellite Constellation is composed of couple of shell groups of satellites.
2. Same shell group of satellite will have the same altitude and inclination.
3. The total N orbit planes in the same shell group of satellites will be evenly distributed by the same interval of Longitude of the ascending node (Ω). The interval equals to $(360 \text{ degree} / N)$. As a result, all orbit planes in the same shell group will effectively form a shell to cover earth (there will be a coverage hole for the shell if the inclination angle is less than 90 degree).
4. Each orbit plane in the same shell group will have the same number of satellites, all satellites in the same orbit plane will be evenly distributed angularly in the orbit plane.
5. All satellites in the same shell group are moving in the same circular direction within the same orbit plane. As a result, at any location on earth, we can see there will have two group of satellites moving on the opposite direction. One group moves from south to north, and another group moves from north to south. Section 5.5 has more details.

4.3. Communication between Satellites by ISL

When ISL is used for the communication between satellites, each satellite will have a fixed number of links to connect to its neighbor. Due to the cost of ISL and the constraints of power supply on satellite, the number of ISL is normally limited to connect to its closest neighbors. In 3D space, each satellite may have six types of adjacent satellites, each type represents one direction. The number of adjacent neighbors in one direction is dependent on the number of deployment of ISL device on satellites, for example, the laser transmitter and receiver for ISLL. Figure 1 illustrates satellite S0 and its adjacent neighbors.

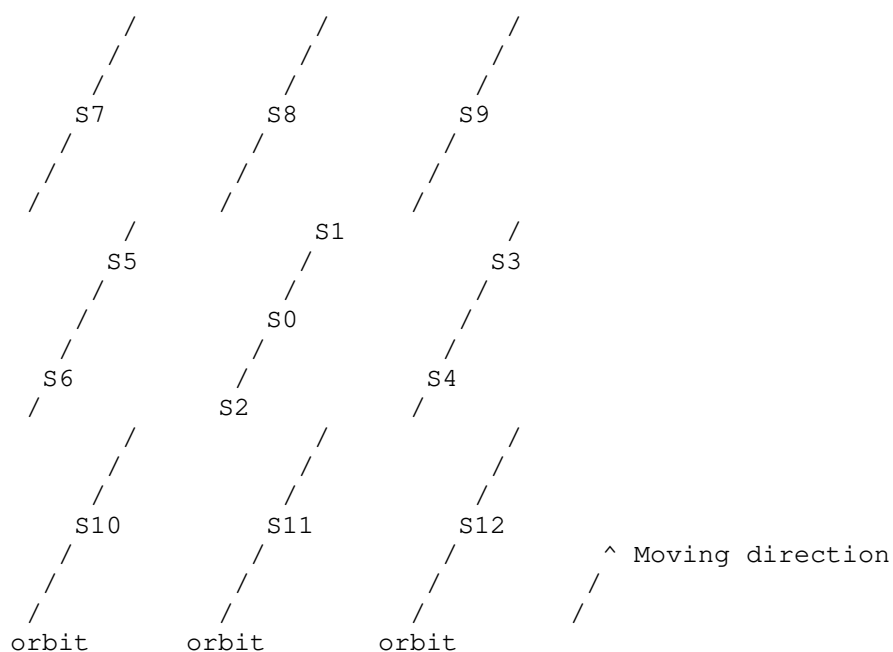


Figure 1: Satellite S0 and its adjacent neighbors

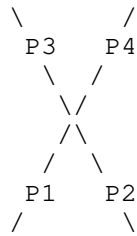
All adjacent satellites of S0 in Figure 1 are listed below:

1. The front adjacent satellite S1 that is on the same orbit plane as S0.
2. The back adjacent satellite S2 that is on the same orbit plane as S0
3. The right adjacent satellites S3 and S4 that are on the right orbit plane of S0

4. The left adjacent satellites S5 and S6 that are on the left orbit plane of S0
5. The above adjacent satellites S7 to S9 that are on the above orbit plane of S0
6. The below adjacent satellite S10 to S12 that are on the below orbit of plane S0

The relative position of adjacent satellites will directly determine the quality of ISL and communication. From the analysis in [I-D.lhan-problems-requirements-satellite-net], The speed of satellite is only related to the altitude of the satellite (on circular orbit), all satellites with a same altitude will move with the same speed. So, in above adjacent satellites, some adjacent satellite's relative positions are steady and the ISL can be alive without interruption caused by movement. Some adjacent satellites relative positions are changing quickly, the ISL may be down since the distance may become out of reach for the laser of ISL, or the quick changed positions of two satellite make the tracking of laser too hard. Below are details:

- * The relative position of satellites in the same orbit plane will be the steadiest.
- * The relative position of satellites in the direct neighbor orbit planes in the same shell group and moving in the same direction will be steady at equator area, but will be changing when two orbits meet on the polar area. Whether the link status will be flipping depends on the tracking technology and the range of laser pointing angle of ISL. See Figure 2.
- * The relative position of satellites in the neighbor orbit planes in the same shell group but moving in the different direction will not be steady at all times. More details are explained in Figure 8
- * The relative position of satellites in the neighbor orbit planes in the different shell group will be dependent on the difference of altitude and inclination. This has been analyzed in [I-D.lhan-problems-requirements-satellite-net].



- * Two satellites S1 and S2 are at position P1 and P2 at time T1
- * S1's right facing ISL connected to S2's left facing ISL
- * S1 and S2 move to the position P4 and P3 at time T2
- * S1's left facing ISL connected to S2's right facing ISL
- * So, if the range of laser pointing angle is 360 degree and tracking technology supports, the ISL will not be flipping after passing polar area; Otherwise, the link will be flipping

Figure 2: Satellite's Position and ISL Change at Polar Area

5. Addressing of Satellite

When ISL is deployed in satellite constellation, all satellites in the constellation can form a network like the wired network on ground. Due to the big number of satellites in a constellation, the network could be either L2 or L3. The document proposes to use L3 network for better scalability.

When satellites form a L3 network, it is expected that IP address is needed for each satellite and its ISLs.

While the traditional IP address can still be used for satellite network, the document proposes an alternative new method for satellite's addressing system. The new addressing system can indicate a satellite's orbit info such as shell group index, orbit plane index and satellite index. This will make the adjacent satellite identification for link status easier and benefit the routing algorithms.

5.1. Indexes of Satellite

As described in Section 4.2, one satellite has three important orbit related information as described below.

1. Index for the shell group of satellites in a satellite constellation
2. Index for the orbit plane in a shell group of satellites

3. Index for the satellite in an orbit plane

It should be noted that for all type of indexes, it is up to the owner to assign the index number. There is no rule for which one should be assigned with which number. The only important rule is that all index number should be in sequential to reflect its relative order and position with others. Below is an example of assignment rules:

1. The 1st satellite launched in an orbit plane can be assigned for the 1st satellite index (0), the incremental direction of the satellite index in the same orbit plane is the incremental direction of "Argument of periapsis (ω)"
2. The 1st orbit plane established can be assigned for the 1st orbit plane index (0), the incremental direction of the orbit plane index is the incremental direction of "Longitude of the ascending node (Ω)".
3. The shell group of satellites with the lowest altitude can be assigned for the 1st shell group index (0), the incremental direction of shell group index is the incremental direction of altitude.

Figure 3 and Figure 4 illustrate three types of indexes for satellite.

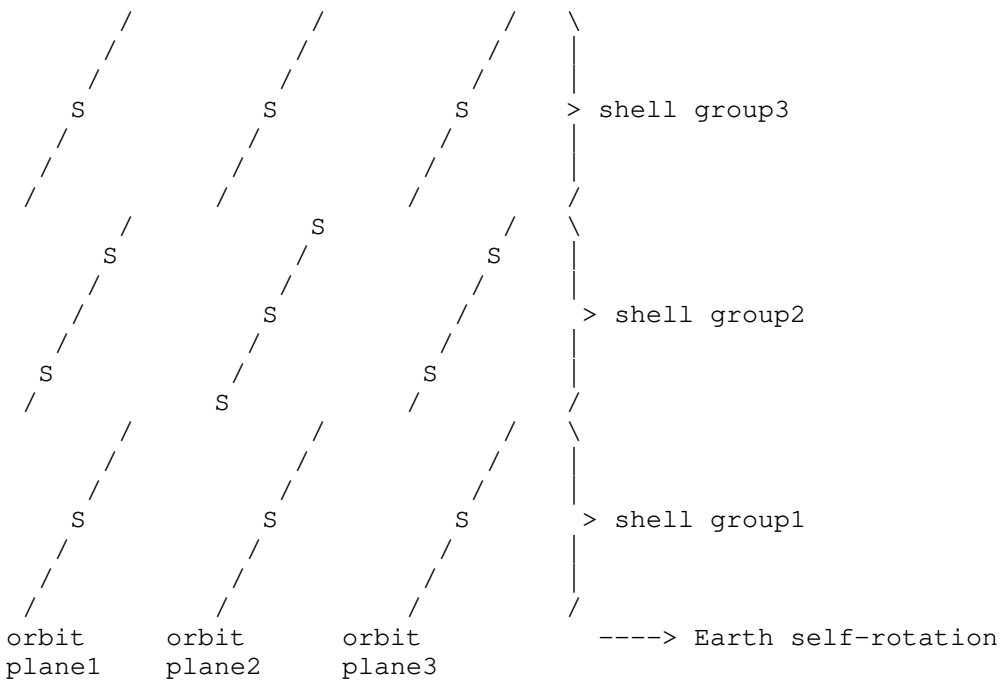


Figure 3: Shell Group and Orbit Plane Indexes for Satellites

Shell Group and Orbit Plane Indexes for Satellites

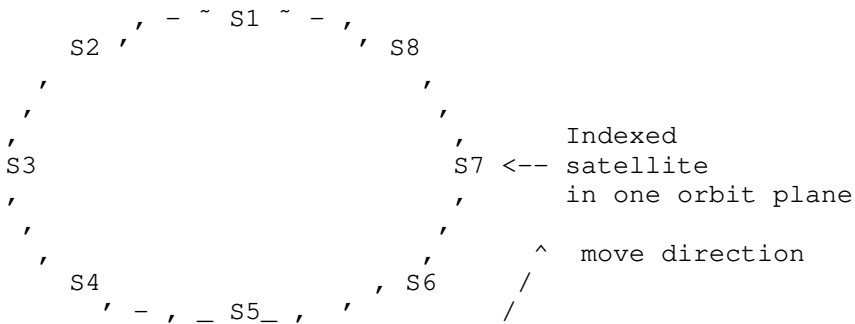


Figure 4

Three type of Index for satellites

5.2. The Range of Satellite Indexes

The ranges of different satellite indexes will determine the range the dedicated field for semantic address. The maximum indexes depend on the number of shell group, orbit plane and satellite per orbit plane. The number of orbit plane and satellite per orbit plane have relationship with the coverage of a satellite constellation. There are minimum numbers required to cover earth.

[I-D.lhan-problems-requirements-satellite-net] has given the detailed math to estimate the minimal number required to cover the earth. There are two key parameters that determine the minimal number of satellite required. One is the elevation angle, another is the altitude. SpaceLink has proposed two elevation angles, 25 and 35 degrees [SpaceX-Non-GEO]. The lowest LEO altitude can be 160km according to [Lowest-LEO-ESA]. The Table 1 and Table 2 illustrate the estimation for different altitude (As), the coverage radius (Rc), the minimal required number of orbit planes (No) and satellite per orbit plane (Ns). The elevation angle is 25 degree and 35 degrees respectively.

Parameters	VLEO1	VLEO2	LEO1	LEO2	LEO3	LEO4	LEO5
As (km)	160	300	600	900	1200	1500	2000
Rc (km)	318	562	1009	1382	1702	1981	2379
Ns	73	42	23	17	14	12	10
No	85	48	27	20	16	14	12

Table 1: Satellite coverage (Rc), minimal number of orbit plane (No) and satellite (Ns) per orbit plane for different LEO/VLEOs, Elevation angle = 25 degree

Parameters	VLEO1	VLEO2	LEO1	LEO2	LEO3	LEO4	LEO5
As (km)	160	300	600	900	1200	1500	2000
Rc (km)	218	392	726	1015	1271	1498	1828
Ns	107	59	32	23	19	16	13
No	123	69	37	27	22	18	15

Table 2: Satellite coverage (Rc), minimal number of orbit plane (No) and satellite (Ns) per orbit for different LEO/VLEOs, Elevation angle = 35 degree

The real deployment may be different as above analysis. Normally, more satellites and orbit planes are used to provide better coverage. So far, there are only two proposals available, one is StarLink, another is from China Constellation. For proposals of [StarLink], there are 7 shell groups, the number of orbit plane and satellites per orbit plane in all shell groups are 72 and 58; For proposals of [China-constellation], there are 7 shell groups, the number of orbit plane and satellites per orbit plane in all shell groups are 60 and 60;

It should be noted that some technical parameters, such as the inclination and altitude of orbit planes, in above proposals may be changed during the long-time deployment period, but the total numbers for indexes normally do not change.

From the above analysis, to be conservative, it is safe to conclude that the range of all three satellite indexes are less than 256, or 8-bit number.

5.3. Other Info for satellite addressing

In addition to three satellite indexes described in Section 5.1, other information is also important and can also be embedded into satellite address:

1. The company or country code, or the owner code. In the future, there may have multiple satellite constellations on the sky from different organizations, and the inter-constellation communication may become as normal that is similar to the network on the ground. This code will be useful to distinguish different satellite constellation and make the inter-constellation communication possible. One satellite constellation will have

one code assigned by international regulator (IANA or ITU). Considering the limit of LEO orbits and the cost of satellite constellations, the total number of satellite constellation is very limited. So, the size of code is limited.

2. The Interface Index. This index is to identify the ISL or ISLL for a satellite. As described in Section 4.3, the total number of ISL is limited. So, the size of interface index is also limited.

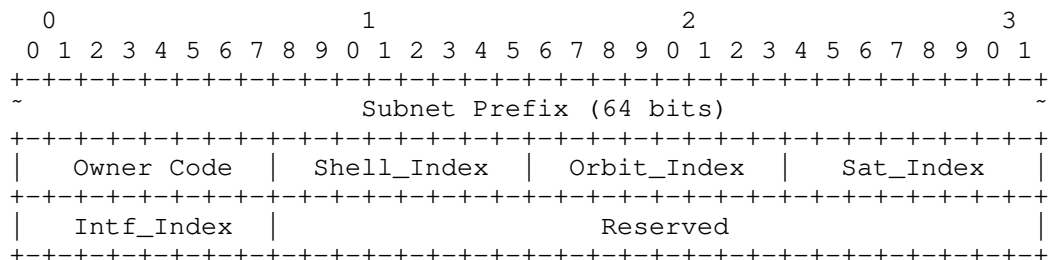
5.4. Encoding of Satellite Semantic Address

The encoding for satellite semantic address is dependent on what routing and switching (L2 or L3 solution) technologies are used for satellite networking, and finally dependent on the decision of IETF community.

Follows are some initial proposals:

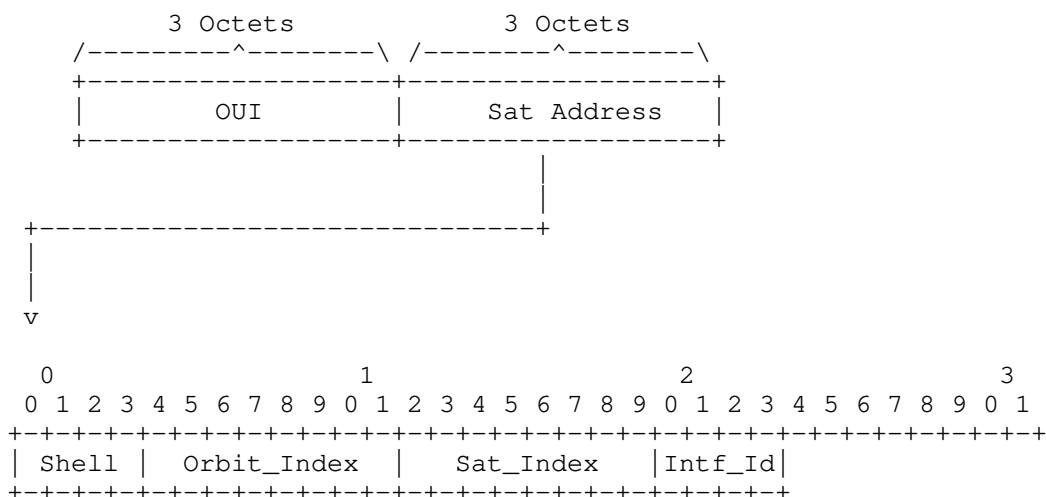
1. When satellite network is using L3 solution, the satellite semantic address is encoded as the interface identifier (i.e., the rightmost 64 bits) of the IPv6 address for IPv6. Figure 5 shows the format of IPv6 Satellite Address.
2. When satellite network is using L2 solution, the satellite semantic address can be embedded into the field of "Network Interface Controller (NIC) Specific" in MAC address [IEEE-MAC-Address]. But due to shorter space for NIC, the "Index for the shell group" and "Index for Interface" will only have 4-bit. This is illustrated in Figure 6. This encoded MAC address can also be used for L3 solution where the interface MAC may be also needed to be configured for each ISL.
3. Recently, some works suggested to use Length Variable IP address for routing and switching [Length-Variable-IP] or use flexible IP address [I-D.jia-flex-ip-address-structure] or shorter IP address [I-D.li-native-short-addresses] to solve some specific problems that regular IPv6 is not very suitable. Satellite network also belongs to such specific network. Due to the resource and cost constraints and requirement for radiation hardened electronic components, the ISL speed, on-board processor and memory are limited in performance, power consumption and capacity compared with network devices on ground. So, using IPv6 directly in satellite network is not an optimal solution because IPv6 header size is too long for such small network. From above analysis, 32-bit to 64-bit length of IP address is enough for satellite networking. Using 128-bit IPv6 will consume more resource especially the ISL bandwidth, processing power and memory, etc.

If shorter than 128-bit IP address is accepted as IETF work, the satellite semantic address can be categorized as a similar use case. Figure 7 illustrates a 32-bit Semantic Satellite Address format. The final coding for the shorter IP address can be decided by the community. How to use the 32-bit Semantic Satellite address can be addressed later on in different document.



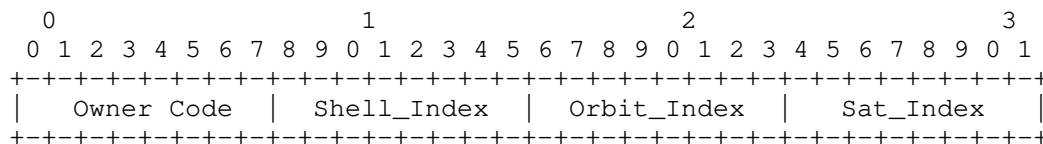
Owner Code: Identifier for the owner of the constellation
 Shell_Index: Index for the shell group of satellite in a satellite constellation
 Orbit_Index: Index for the orbit plane in a shell group of satellite
 Sat_Index: Index for the satellite in an orbit plane
 Intf_Index: Index for interface on a satellite
 Reserved: 24-bits reserved

Figure 5: The IPv6 Satellite Address



```
OUI: Organizationally Unique Identifier assigned by IEEE
Shell: 4-bit Index for the shell group of satellite in a satellite
      constellation
Orbit_Index: Index for the orbit plane in the group of satellite
Sat_Index: Index for the satellite in the orbit plane
Intf Id: 4-bit Index for interface on a satellite
```

Figure 6: The MAC Satellite Address



```
Owner Code: Identifier for the owner of the constellation
Shell_Index: Index for the shell group of satellite in a satellite
              constellation
Orbit_Index: Index for the orbit plane in a shell group of satellite
Sat_Index: Index for the satellite in an orbit plane
```

Figure 7: The 32-bit Semantic Satellite Address

5.5. Link Identification by Satellite Semantic Address

Using above satellite semantic addressing scheme, to identify steady and un-steady links is as simple as below:

Assuming:

1. The total number of satellites per orbit plane is M
2. The total number of orbit planes per shell group is N.
3. Two satellites have:
 - * Satellite Indexes as: Sat1_Index, Sat2_Index
 - * Orbit plane Indexes as: Orbit1_Index, Orbit2_Index
 - * Shell group Indexes as: Shell1_Index, Shell2_Index

Steady links:

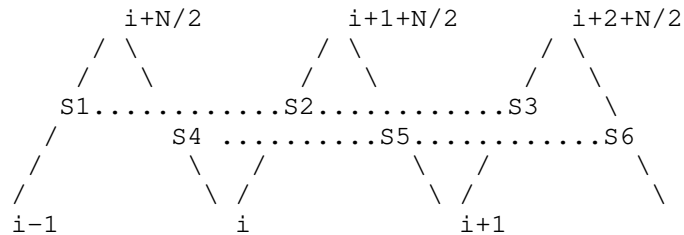
1. The links between adjacent satellites on the same orbit plane, or, the satellite indexes satisfy:
 - * $\text{Sat2_Index} = \text{Sat1_Index} + 1$, when $\text{Sat1_Index} < M-1$; $\text{Sat2_Index} = 0$, when $\text{Sat1_Index} = M-1$; and
 - * $\text{Orbit1_Index} = \text{Orbit2_Index}$, $\text{Shell1_Index} = \text{Shell2_Index}$.
2. The links between satellites on adjacent orbit planes on the same altitude. and two satellites are moving to the same direction, or, the satellite indexes satisfy:
 - * $\text{Orbit2_Index} = \text{Orbit1_Index} + 1$, when $\text{Orbit1_Index} < N-1$; $\text{Orbit2_Index} = 0$, when $\text{Orbit1_Index} = N-1$; and
 - * $\text{Shell1_Index} = \text{Shell2_Index}$.
 - * Sat1_Index and Sat2_Index may be equal or have difference, depend on how the link is established.

Un-Steady links:

1. The links between satellite and ground stations.
2. The links between satellites on adjacent orbit planes on the same altitude. Two satellites are moving to the different direction. Or, the satellite indexes do not satisfy conditions described in above #2 for Steady links.
3. The links between satellites on adjacent orbit planes on different altitude. Or, the satellite indexes satisfy:

* Shell1_Index != Shell2_Index.

Figure 8 illustrates the links for adjacent orbit planes (#2 for Steady Link and Un-steady Link above). From the figure, it can be noticed that some links may have shorter distance than steady link, but they are unsteady. For example, the links between S1 and S4; S4 and S2; S2 and S5, etc.



- * The total number of orbit planes are N
- * The number (i-1, i, i+1,...) represents the Orbit index
- * The bottom numbers (i-1, i, i+1) are for orbit planes on which satellites (S1, S2, S3) are moving from bottom to up.
- * The top numbers (i+N/2, i+1+N/2, i+2+N/2) are for orbit planes on which satellites (S4, S5, S6) are moving from up to bottom.
- * Dot lines are the steady links

Figure 8: The links between satellites on adjacent orbit planes

6. IANA Considerations

This memo may include request to IANA for owner code, see Section 5.4.

7. Contributors

8. Acknowledgements

9. References

9.1. Normative References

- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.

- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <<https://www.rfc-editor.org/info/rfc2328>>.
- [RFC7142] Shand, M. and L. Ginsberg, "Reclassification of RFC 1142 to Historic", RFC 7142, DOI 10.17487/RFC7142, February 2014, <<https://www.rfc-editor.org/info/rfc7142>>.
- [RFC2453] Malkin, G., "RIP Version 2", STD 56, RFC 2453, DOI 10.17487/RFC2453, November 1998, <<https://www.rfc-editor.org/info/rfc2453>>.
- [RFC7868] Savage, D., Ng, J., Moore, S., Slice, D., Paluch, P., and R. White, "Cisco's Enhanced Interior Gateway Routing Protocol (EIGRP)", RFC 7868, DOI 10.17487/RFC7868, May 2016, <<https://www.rfc-editor.org/info/rfc7868>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

9.2. Informative References

- [I-D.lhan-problems-requirements-satellite-net]
Han, L., Li, R., Retana, A., Chen, M., Su, L., and N. Wang, "Problems and Requirements of Satellite Constellation for Internet", Work in Progress, Internet-Draft, draft-lhan-problems-requirements-satellite-net-02, 13 February 2022, <<https://datatracker.ietf.org/doc/html/draft-lhan-problems-requirements-satellite-net-02>>.
- [I-D.jia-flex-ip-address-structure]
Jia, Y., Chen, Z., and S. Jiang, "Flexible IP: An Adaptable IP Address Structure", Work in Progress, Internet-Draft, draft-jia-flex-ip-address-structure-00, 31 October 2020, <<https://datatracker.ietf.org/doc/html/draft-jia-flex-ip-address-structure-00>>.
- [I-D.li-native-short-addresses]
Li, G., Jiang, S., and D. E. 3rd, "Native Short Addresses for the Internet Edge", Work in Progress, Internet-Draft, draft-li-native-short-addresses-01, 25 May 2021, <<https://datatracker.ietf.org/doc/html/draft-li-native-short-addresses-01>>.

[Routing-for-LEO]

E. Ekici, I. F. Akyildiz and M. D. Bender, "Datagram routing algorithm for LEO satellite networks," Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064), 2000, pp. 500-508 vol.2, doi: 10.1109/INFCOM.2000.832223.", <<https://ieeexplore.ieee.org/document/832223>>.

[Length-Variable-IP]

Shoushou Ren, Delei Yu, Guangpeng Li, Shihui Hu, Ye Tian, Xiangyang Gong, Robert Moskowitz, "Routing and Addressing with Length Variable IP Address," NEAT'19: Proceedings of the ACM SIGCOMM 2019 Workshop on Networking for Emerging Applications and Technologies, August 2019", <<https://doi.org/10.1145/3341558.3342204>>.

[IEEE-MAC-Address]

"IEEE Std 802-2001 (PDF). The Institute of Electrical and Electronics Engineers, Inc. (IEEE). 2002-02-07. p. 19. ISBN 978-0-7381-2941-9. Retrieved 2011-09-08.", <<https://standards.ieee.org/getieee802/download/802-2001.pdf>>.

[Lowest-LEO-ESA]

"Lowest LEO by ESA", <[https://www.esa.int/ESA_Multimedia/Images/2020/03/Low_Earth_orbit#:~:text=A%20low%20Earth%20orbit%20\(LEO,very%20far%20above%20Earth's%20surface.>](https://www.esa.int/ESA_Multimedia/Images/2020/03/Low_Earth_orbit#:~:text=A%20low%20Earth%20orbit%20(LEO,very%20far%20above%20Earth's%20surface.>)>.

[KeplerianElement]

"Keplerian elements", <https://en.wikipedia.org/wiki/Orbital_elements>.

[OSI-Model]

"OSI Model", <https://en.wikipedia.org/wiki/OSI_model>.

[StarLink] "Star Link", <<https://en.wikipedia.org/wiki/Starlink>>.

[China-constellation]

"China Constellation", <<https://www.itu.int/ITU-R/space/asreceived/Publication/DisplayPublication/23706>>.

[SpaceX-Non-GEO]

"FCC report: SPACEX V-BAND NON-GEOSTATIONARY SATELLITE SYSTEM", <<https://fcc.report/IBFS/SAT-LOA-20170301-00027/1190019.pdf>>.

Appendix A. Change Log

- * Initial version, 10/19/2021
- * 01 version, 02/28/2022

Authors' Addresses

Lin Han (editor)
Futurewei Technologies, Inc.
2330 Central Express Way
Santa Clara, CA 95050,
United States of America
Email: lhan@futurewei.com

Richard Li
Futurewei Technologies, Inc.
2330 Central Express Way
Santa Clara, CA 95050,
United States of America
Email: rli@futurewei.com

Alvaro Retana
Futurewei Technologies, Inc.
2330 Central Express Way
Santa Clara, CA 95050,
United States of America
Email: alvaro.retana@futurewei.com

Meiling Chen
China Mobile
32, Xuanwumen West
BeiJing 100053
China
Email: chenmeiling@chinamobile.com

Ning Wang
University of Surrey
Guildford
Surrey, GU2 7XH
United Kingdom
Email: n.wang@surrey.ac.uk

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 8, 2022

Z. Li
S. Peng
Huawei Technologies
D. Voyer
Bell Canada
C. Li
China Telecom
P. Liu
China Mobile
C. Cao
China Unicom
G. Mishra
Verizon Inc.
March 7, 2022

Application-aware Networking (APN) Framework
draft-li-apn-framework-05

Abstract

A multitude of applications are carried over the network, which have varying needs for network bandwidth, latency, jitter, and packet loss, etc. Some new emerging applications have very demanding performance requirements. However, in current networks, the network and applications are decoupled, that is, the network is not aware of the applications' requirements in a fine granularity. Therefore, it is difficult to provide truly fine-granularity traffic operations for the applications and guarantee their SLA requirements.

This document proposes a new framework, named Application-aware Networking (APN), where application-aware information (i.e. APN attribute) including APN identification (ID) and/or APN parameters (e.g. network performance requirements) is encapsulated at network edge devices and carried in packets traversing an APN domain in order to facilitate service provisioning, perform fine-granularity traffic steering and network resource adjustment.

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 September 8, 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Requirements Language	4
3. Terminology	4
4. APN Framework and Key Components	4
5. APN Requirements	6
5.1. APN Attribute Conveying Requirements	6
5.1.1. Protocol Extensions Requirements	8
5.2. APN attribute Handling Requirements	9
5.2.1. Fine granular SLA Guarantee	9
5.2.2. Fine granular network slicing	10
5.2.3. Fine granular deterministic networking	11
5.2.4. Fine granular service function chaining	11
5.2.5. Fine granular network measurement	12
6. Illustration	12
6.1. Example use case description	12
6.2. User Group and Application Group Design	13
6.3. Derive the User Group and User Group at APN Edge	15
6.4. Access Right Check at the edge of the backbone network	15
6.5. SLA Guarantee in the backbone network	16
6.5.1. Network Measurement	16
6.5.2. Traffic Steering	17
7. Benefits	17
8. IANA Considerations	18
9. Security Considerations	18

10. Acknowledgements	19
11. Co-authors	19
12. Contributors	20
13. References	21
13.1. Normative References	21
13.2. Informative References	21
Authors' Addresses	22

1. Introduction

A multitude of applications are carried over the network, which have varying needs for network bandwidth, latency, jitter, and packet loss, etc. Some applications such as online gaming and live video streaming have very demanding network requirements and therefore require special treatment in the network. However, in current networks, the network and applications are decoupled, that is, the network is not aware of the applications' requirements in a fine granularity. Therefore, it is difficult to provide truly fine-granularity traffic operations for the applications and guarantee their SLA requirements accordingly.

[I-D.li-apn-problem-statement-usecases] describes the challenges of traditional differentiated service provisioning methods, such as five tuples used for ACL/PBR causing coarse granularity as well as orchestration and SDN-based solution causing long control loops.

This document proposes a new framework, named Application-aware Networking (APN), where application-aware information (APN attribute) including application-aware identification (APN ID) and application-aware parameters (APN Parameters), is encapsulated at network edge devices and carried along with the encapsulation of the tunnel that is used by the packet to traverse the APN domain. The APN attribute will facilitate service provisioning and provide fine-granularity services in the APN domain.

The APN attribute is acquired based on the existing information in the packet header (i.e. source and destination addresses, incoming L2 (or) MPLS encapsulation, incoming physical/virtual port information, the other fields of the 5-tuple if they are not encrypted) at the edge devices of the APN domain, added to the data packets along with the tunnel encapsulation, delivered within the network, and removed when the packets leave the domain together with the tunnel encapsulation.

APN aims to apply various policies in different nodes along a network path onto a traffic flow altogether, for example, at the headend to steer into corresponding path, at the midpoint to collect corresponding performance measurement data, and at the service function to execute particular policies.

APN is only applied to an edge-to-edge tunnel encapsulation within a limited trusted domain. It means that the source and destination addresses of the packet are the endpoints of the tunnel (i.e. the domain edges), and nothing about the payload source and destination can be deduced, which substantially reduces the privacy concerns. Typically, an APN domain is defined as a Network Operator controlled limited domain (see Figure 1), in which MPLS, VXLAN, SR/SRv6 and other tunnel technologies are adopted to provide network services.

2. Requirements Language

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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminology

ACL: Access Control List

APN: Application-aware Networking

APN6: Application-aware Networking for IPv6/SRv6

LB: Load Balancing

MPLS: Multiprotocol Label Switching

PBR: Policy Based Routing

QoE: Quality of Experience

SDN: Software Defined Networking

SLA: Service Level Agreement

SR: Segment Routing

SR-MPLS: Segment Routing over MPLS dataplane

SRv6: Segment Routing over IPv6 dataplane

4. APN Framework and Key Components

The APN framework is shown in Figure 1. The key components include App-aware Edge Device (APN-Edge), App-aware-process Head-End (APN-

Head), App-aware-process Mid-Point (APN-Midpoint), and App-aware-process End-Point (APN-Endpoint).

Packets carry application characteristic information (i.e. APN attribute) which includes the following information:

- o Application-aware identification (APN ID): identifies the set of attributes, indicating that all packets belonging to the same flow will be given the same treatment by the network. ;
- o Application-aware parameters (APN parameters): The typical application-aware parameters are the network performance requirement parameters including bandwidth, delay, delay variation, packet loss ratio, etc.

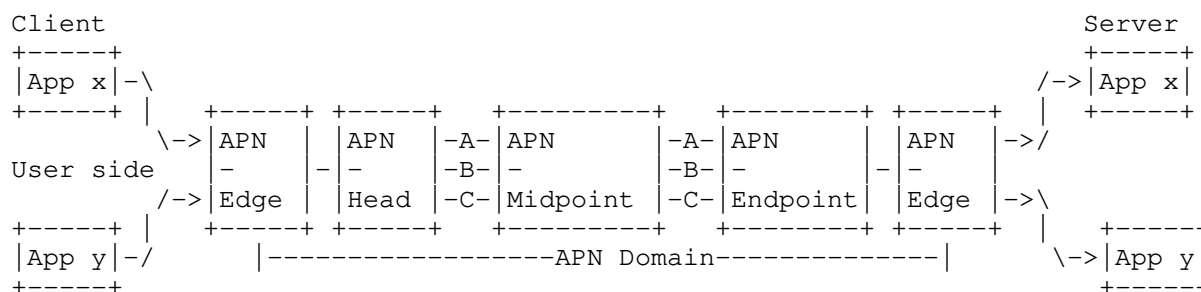


Figure 1: Framework and Key Components

The key components are introduced as follows.

- o App-aware Edge Device (APN-Edge): this network device receives packets from applications and obtains the APN attribute based on the configuration on this device according to the existing information in the packet header, such as 5-tuple, VLAN or double VLAN tagging (C-VLAN and S-VLAN). The APN-Edge device adds the APN attribute in the tunnel encapsulation. The packets carrying the APN attribute will be sent to the APN-Head, and the APN attribute will be used to apply various policies in different nodes along the network path onto the traffic flow, e.g., at the headend to steer into corresponding path satisfying SLAs, at the midpoint to collect corresponding performance measurement data, at the service function to execute particular policies. When the packets leave the APN domain, the APN attribute will be removed together with the tunnel encapsulation.
- o App-aware-process Head-End (APN-Head): This network device receives packets from the APN-Edge, obtains the APN attribute, and

initiates the corresponding process. Generally, in order to satisfy different SLA requirements, a set of paths, tunnels or SR policies, are set up between the APN-Head and the APN-Endpoint. These multiple parallel paths have different SLA guarantees. The APN-Head maintains the matching relationship between the APN attribute and the paths between the APN-Head and the APN-Endpoint. The APN-Head determines the path between the APN-Head and the APN-Endpoint according to the APN attribute carried in the packets and the matching relationship with it, which satisfies the service requirements of the applications. The APN-Head forwards the packets along the path. The APN attribute conveyed by the packet received from the APN-Edge can also be copied or be mapped to the outgoing packet header.

- o App-aware-process Mid-Point (APN-Midpoint): the APN-Midpoint provides the path service and enforces various policies according to the APN attribute carried in the packets. The APN-Midpoint may also adjust the resource locally to guarantee the service requirements depending on a specific policy and the APN attribute conveyed by the packet. Policy definitions and mechanisms are out of the scope of this document.
- o App-aware-process End-Point (APN-Endpoint): the process of the specific service path will end at the APN-Endpoint. If the outer tunnel header for the path between the APN-Head and the APN-Endpoint exists, it will be removed by the APN-Endpoint. If the APN attribute is copied or mapped to the outer tunnel header by the APN-Head, it will also be removed along with the outer tunnel header.

Note that in the actual implementation, the APN-Edge can co-exist with the APN-Head or APN-Endpoint, that is, one network device can implement the functionalities of both APN-Edge and the APN-Head/APN-Endpoint.

5. APN Requirements

This section specifies the requirements for supporting the APN framework, including the requirements for conveying and handling the APN attribute.

5.1. APN Attribute Conveying Requirements

The APN attribute consists of APN ID and APN parameters.

APN ID includes the following identifiers (IDs),

- o Application Group ID: identifies an application group of the traffic.
- o User Group ID: identifies the user group of the traffic.

APN ID can be acquired through different ways. In the APN work it MUST be acquired according to the existing available information in the packet header without inspection into the payload.

The different combinations of the IDs can be used to provide different granularity of the service provisioning and SLA guarantee for the traffic.

The APN parameters are the network performance requirement parameters. The network service requirement can include the following parameters:

- o Bandwidth: the bandwidth requirement
- o Latency: the latency requirement
- o Packet loss ratio: the packet loss ratio requirement
- o Jitter: the jitter requirement

The different combinations of the parameters are for further expressing the more detailed service requirements, conveyed together with the APN ID, which can be used to match to appropriate tunnels/SR Policies and queues that can satisfy these service requirements.

APN attribute MUST be encapsulated within tunnels in the network layer. The tunnels include but not limit to MPLS, VxLAN, SR-MPLS, and SRv6. It can be extended according to requirements in the future.

[REQ 1a]. APN ID SHOULD include Application Group ID to indicate the application group that the packet belongs to.

[REQ 1b]. APN ID SHOULD include User Group ID to indicate the user group that the packet belongs to.

[REQ 1c]. APN ID MUST include either Application Group ID or User Group ID.

[REQ 1d]. APN ID MUST be acquired from the existing available information of the packet header without interference into the payload.

[REQ 1e]. APN parameters is OPTIONAL.

[REQ 1f]. APN attribute MUST be carried by the outer tunnel encapsulation.

[REQ 1g]. All the nodes along the path SHOULD be able to process the APN attribute if needed.

[REQ 1h]. The APN attribute is generated by the APN-Edge though local policy.

[REQ 1i]. The APN attribute SHOULD be kept intact when directly copied at the APN-Head and carried in the tunnel encapsulation.

[REQ 1j]. The APN attribute MUST be removed along with the tunnel encapsulation by the APN-Edge when the packets leave the APN domain.

[REQ 1k]. The APN attribute MUST NOT be encrypted when the APN packet is itself encrypted (e.g., the APN tunnel across the APN domain uses IPsec).

5.1.1. Protocol Extensions Requirements

The APN attribute is conveyed with the tunnel encapsulation. There are two typical types of tunnels:

- o MPLS-based tunnel: LDP tunnel, RSVP-TE tunnel, SR-MPLS tunnel or policy, etc.
- o IPv6-based tunnel: IPv6-based VxLAN tunnel, IPv6-based UDP tunnel, IPv6-based GRE tunnel, SRv6 tunnel or policy, etc.

In order to support encapsulation of APN attribute, the MPLS data plane and IPv6 data plane need to be extended.

In order to support acquiring the APN attribute according to the existing available information in the packet header, YANG models should be defined to configure the mapping between the application/user group ID and the existing information in the packet header and configure the corresponding APN attribute for the application/user group. It can also be implemented with protocol extensions such as BGP and PCEP which can advertise the information from the central controller to the APN-Edge.

In addition, in the APN domain, the above-mentioned mapping and applying APN parameters may also be advertised from the APN-Edge/APN-Head to other devices or from the network devices to the central

controller in the APN domain. IGP extensions or BGP-LS extensions should be introduced to achieve the purposes.

[REQ 1-1a] MPLS encapsulation SHOULD be extended to be able to carry the APN attribute for MPLS-based tunnels.

[REQ 1-1b] IPv6 encapsulation SHOULD be extended to be able to carry the APN attribute for IPv6-based tunnels.

[REQ 1-1c] YANG models SHOULD be defined to implement the mapping between the application/user group ID and the existing available information in the packet header and configure the corresponding APN parameters.

[REQ 1-1d] BGP extensions SHOULD be defined to advertise the mapping between the application/user group ID and the existing available information in the packet header and the corresponding APN parameters from the central controller to the APN-Edge in the APN domain.

[REQ 1-1e] PCEP extensions SHOULD be defined to advertise the mapping between the application/user group ID and the existing available information in the packet header and the corresponding APN parameters from the central controller to the APN-Edge in the APN domain.

[REQ 1-1f] IGP extensions SHOULD be defined to advertise the mapping between the application/user group ID and the existing available information in the packet header and the corresponding APN parameters from the APN-Edge to the network devices in the APN domain.

[REQ 1-1g] BGP-LS extensions SHOULD be defined to advertise the mapping between the application/user group ID and the existing available information in the packet header and the corresponding APN parameters from the network devices to the central controller in the APN domain.

5.2. APN attribute Handling Requirements

The APN Head and APN-Midpoint perform matching operation against the APN attribute, that is, to match IDs and/or service requirements to the corresponding network resources such as tunnels/SR policies and queues.

5.2.1. Fine granular SLA Guarantee

In order to achieve better Quality of Experience (QoE) of end users and engage customers, the network needs to be able to provide fine-granularity SLA guarantee [I-D.li-apn-problem-statement-usecases].

[REQ 2-1a]. With the APN attribute, the APN-Head SHOULD be able to steer the traffic to the tunnel/SR policy that satisfies the matching operation.

[REQ 2-1b]. With the APN attribute, the APN-Head SHOULD be able to trigger the setup of the tunnel/SR policy that satisfies the matching operation.

[REQ 2-1c]. With the APN attribute, the APN-Head and APN-Midpoint SHOULD be able to steer the traffic to the queue that satisfies the matching operation.

[REQ 2-1d]. With the APN attribute, the APN-Head and APN-Midpoint SHOULD be able to trigger the configuration of the queue that satisfies the matching operation.

[REQ 2-1e]. If the tunnels are used to satisfy the performance requirements, the APN-Head SHOULD be able to copy or map the APN attribute conveyed by the packet received from the APN-Edge to the outer tunnel header.

[REQ 2-1f]. If the tunnels are used to satisfy the performance requirements and the APN attribute are conveyed along with the outer tunnel, the APN-Endpoint MUST remove the APN attribute along with the outer tunnel.

5.2.2. Fine granular network slicing

Network slicing provides ways to partition the network infrastructure in either control plane or data plane into multiple network slices that are running in parallel. The resources on each node need to be associated to corresponding slices.

APN is to help the operator of a network to steer some of the traffic tagged with an APN attribute to a certain network slice based on the SLA agreement with its customer.

[REQ 2-2a]. With the APN attribute, the APN-Head SHOULD be able to steer the traffic to the slice that satisfies the matching operation.

[REQ 2-2b]. With the APN attribute, the APN-Midpoint SHOULD be able to associate the traffic to the resources in the slice that satisfies the matching operation.

5.2.3. Fine granular deterministic networking

Along the path each node needs to provide guaranteed bandwidth, bounded latency, and other properties relevant to the transport of time-sensitive data for the Detnet flows that coexist with the best-effort traffic.

APN is to help the operator of a network to steer some of the traffic tagged with an APN attribute to a certain deterministic path based on the SLA agreement with its customer.

[REQ 2-3a]. With the APN attribute, the APN-Head SHOULD be able to steer the traffic to the appropriate path that satisfies the matching operation.

[REQ 2-3b]. With the APN attribute, the APN-Head SHOULD be able to trigger the setup of the appropriate path that satisfies the matching operation for the Detnet flows.

[REQ 2-3c]. With the APN attribute, the APN-Midpoint SHOULD be able to associate the traffic to the resources along the path that satisfies the performance guarantee.

[REQ 2-3d]. With the APN attribute, the APN-Midpoint SHOULD be able to reserve the resources for the Detnet flows along the path that satisfies the performance guarantee.

5.2.4. Fine granular service function chaining

The end-to-end service delivery often needs to go through various service functions, including traditional network service functions such as firewalls, LB as well as new application-specific functions, both physical and virtual. SFC is applicable to both fixed and mobile networks as well as data center networks.

APN is to help the operator of a network to steer some of the traffic tagged with an APN attribute to a certain service function chain based on the SLA agreement with its customer. On each service function along the service function chain, the policy can be enforced based on the APN attribute in the outer header.

[REQ 2-4a]. With the APN attribute, the App-aware-process devices SHOULD be able to steer the traffic to the appropriate service function.

[REQ 2-4b]. The App-aware-process devices including VAS SHOULD be able to process the APN attribute carried in the packets.

5.2.5. Fine granular network measurement

Network measurement can be used for verifying whether the network performance requirements have been satisfied, as well as locating silent failure and predicting QoE satisfaction, which enables real-time SLA awareness/proactive OAM and potential resource adjustments.

APN is to help the operator of a network to trigger performance measurement for the traffic tagged with an APN attribute based on its customer' consent.

[REQ 2-5a]. The App-aware-process devices SHOULD be able to perform IOAM based on the APN attribute.

[REQ 2-5b]. The network measurement results can be reported based on the APN attribute and verify whether the performance requirements are satisfied.

6. Illustration

In order to better clarify what APN can enable with the introduced APN attribute compared to the existing network without APN, we illustrate how APN works through an example use case, which is also a typical network service being provisioned nowadays, i.e. the Cloud Leased Line service. In order to make the tunnel description much easier to understand, we use the recent technology in IETF, i.e. SRv6.

6.1. Example use case description

We take the "SRv6-based Cloud Leased Line Service" as an illustrative example to show how APN is needed and can be beneficial.

Enterprises usually buy Cloud Leased Line Service to interconnect their local sites to Cloud. Generally, the Cloud Leased Line Service needs to go across multiple domains which are owned by the same operator and can be controlled by multiple controllers and an orchestrator/super-controller.

Due to management reasons, the network information in the intermediate domain cannot be advertised to other domains, so the ingress node cannot set up an appropriate E2E path. In that case, the intermediate domain is treated as a black box, and no fine grain traffic steering and other services can be provisioned.

The example of the network to provide the cloud leased lined service reference diagram is shown as the following figure. The network is composed by three network domains including the two metro networks in

the City A and City B and the backbone network which connects the two metro networks. The cloud leased line services is provided to the specific enterprise whose branches located in different cities need to access the cloud-based service located in the City B.

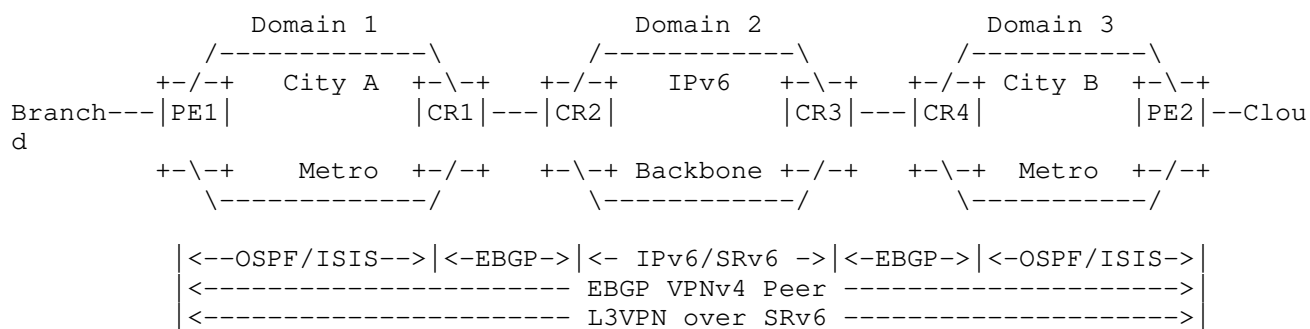


Figure 2. Reference diagram for the example use case illustration

6.2. User Group and Application Group Design

The user groups can be designed as follows:

	User Group
Enterprise A/Branch 1/Office Users	001001001
Enterprise A/Branch 1/R&D Users	001001002
Enterprise A/Branch 1/IT Users	001001003
Enterprise A/Branch 1/VIP Users	001001004
Enterprise A/Branch 2/Office Users	001002001
Enterprise A/Branch 2/R&D Users	001002002
Enterprise A/Branch 2/IT Users	001002003
Enterprise A/Branch 2/VIP Users	001002004
Enterprise A/Branch 3/Office Users	001003001
Enterprise A/Branch 3/R&D Users	001003002
Enterprise A/Branch 3/IT Users	001003003
Enterprise A/Branch 3/VIP Users	001003004

In the IP address design, the IPv6 address blocks allocated to the branches are as follows :

	IPv6 Address
Enterprise A/Branch 1/Office Users	2001:DB8:A:11::/56
Enterprise A/Branch 1/R&D Users	2001:DB8:A:12::/56
Enterprise A/Branch 1/IT Users	2001:DB8:A:13::/56
Enterprise A/Branch 1/VIP Users	2001:DB8:A:1D::/56
Enterprise A/Branch 2/Office Users	2001:DB8:A:21::/56
Enterprise A/Branch 2/R&D Users	2001:DB8:A:22::/56
Enterprise A/Branch 2/IT Users	2001:DB8:A:23::/56
Enterprise A/Branch 2/VIP Users	2001:DB8:A:2D::/56
Enterprise A/Branch 3/Office Users	2001:DB8:A:31::/ 56
Enterprise A/Branch 3/R&D Users	2001:DB8:A:32::/56
Enterprise A/Branch 3/IT Users	2001:DB8:A:33::/56
Enterprise A/Branch 3/VIP Users	2001:DB8:A:3D::/56

The application groups provided by the cloud can be designed as follows:

	Application Group
Enterprise A/Office Audio Applications	101001001
Enterprise A/Office Video Applications	101001002
Enterprise A/Office Data Applications	101001003
Enterprise A/R&D Audio Applications	101002001
Enterprise A/R&D Video Applications	101002002
Enterprise A/R&D Data Applications	101002003
Enterprise A/IT Audio Applications	101003001
Enterprise A/IT Video Applications	101003002
Enterprise A/IT Data Applications	101003003

In the address design, the IPv6 address blocks allocated to the applications of Enterprise A in the cloud is 2001:DB8:A1::/48A1::A:/16. The port number can be used to identify different applications.

	IPv6 Address	Port Num
ber		
19	Enterprise A/Office Audio Applications	2001:DB8:A1:A1::/64 1718, 17
61	Enterprise A/Office Video Applications	2001:DB8:A1:A1::/64 5060, 50
19	Enterprise A/Office Data Applications	2001:DB8:A1:A1::/64 21, 80
61	Enterprise A/R&D Audio Applications	2001:DB8:A1:A2::/64 1718, 17
19	Enterprise A/R&D Video Applications	2001:DB8:A1:A2::/64 5060, 50
61	Enterprise A/R&D Data Applications	2001:DB8:A1:A2::/64 21, 80
19	Enterprise A/IT Audio Applications	2001:DB8:A1:A3::/64 1718, 17
61	Enterprise A/IT Video Applications	2001:DB8:A1:A3::/64 5060, 50
	Enterprise A/IT Data Applications	2001:DB8:A1:A3::/64 21, 80

6.3. Derive the User Group and User Group at APN Edge

The cloud leased line service adopts the SRv6-based L3VPN to traverse the network. The following policy can be applied at the APN edges of the City A1:

```
Match:
  VPN1
  Source Address 2001:DB8:A:11::/56
Action
  Set user-group 001001001

Match:
  VPN1
  destination Address 2001:DB8:A1:A1::/64
  destination port 1718,1719
Action
  Set app-group 101001001
```

6.4. Access Right Check at the edge of the backbone network

The following check can be applied at the edge of the IP backbone network:

```
Match:
  user-group 001001001
  app-group 101002001, 101002002, 101002003, 101003001, 101003002, 101003003
Action
  Deny

Match:
  user-group 001001001
  app-group 101001001, 101001002, 101001003
Action
  Permit
```

The policy means that the office users of the branch 1 can only access the office applications.

If the address allocation is changed. For example, one office user of the branch1's IPv6 address is changed to 2001:DB8:A:15::/56 because of the mobile office.

We only need to add the following policy at the APN edge:

```
Match:
  VPN1
  Source Address 2001:DB8:A:15::/56
Action
  Set user-group 001001001
```

The policy in the backbone network which is based on the user group and the application group is not necessary to change.

6.5. SLA Guarantee in the backbone network

Due to management reasons, the network information in the intermediate domain cannot be advertised to other domains, so the ingress node cannot set up an appropriate TE path, the intermediate domain is treated as a black box and no fine grain traffic steering can be performed.

In this case, we consider fine grain traffic steering in Domain 2 on top of the SRv6-based Cloud Leased Line Service for the purpose of SLA Guarantee.

6.5.1. Network Measurement

In order to guarantee SLA for the VIP users, the following network measurement policy can be applied in the backbone network:

```
Match:
  User-group 001001004 application group 101001002
  User-group 001002004 application group 101002002
  User-group 001003004 application group 101003002
Action
  Apply IOAM
```

The policy is to apply the IOAM as the network measurement for the VIP users of the branches to access the video applications. From the above illustration, there is the following observation:

When there is no APN deployed, at CR2, the 5 tuples of the original packets will need to be resolved since they have been encapsulated, and then IOAM can be triggered based on the 5 tuples. This resolution process is costly and consumes a lot of hardware resources. If Domain 3 needs to trigger IOAM, the same resolution process will have to be done at CR4.

When there is APN deployed, at CR1, the APN attribute is tagged. When these packets arrive at CR2, only the APN attribute in the outer header will be read out, based on which the IOAM can be triggered in

Domain 2. That is, no 5-tuple resolution process is needed at CR2 but only checking the APN attribute in the outer header.

6.5.2. Traffic Steering

If the SLA guarantee of the VIP users accessing the video applications does not satisfy the requirements through the network measurement based on the IOAM, the SRv6 policy can be setup. For example, the SRv6 policy 1 which can satisfy the SLA requirement is set up. Then the following policy can be downloaded to the edge:

Match:

User-group 001001004 application group 101001002

User-group 001002004 application group 101002002

User-group 001003004 application group 101003002

Action

Redirect SRv6 Policy 1

The policy is to steer the traffic of the VIP users to the SRv6 policy in the backbone to satisfy the requirement .

From the above illustration, there is the similar observation as the network measurement:

When there is no APN deployed, at CR2, the 5 tuples of the original packets will need to be resolved since they have been encapsulated, and then the traffic can be steered into SRv6 policy 2 based on the 5 tuples. This resolution process is costly and consumes a lot of hardware resources.

When there is APN deployed, at CR1, the APN attribute is tagged. When these packets arrive at CR2, only the APN attribute in the outer header will be read out, based on which the traffic can be steered into SRv6 policy 2 in Domain 2.

7. Benefits

The APN attribute allows the network devices to only look at one easily-accessible field in the outer header, without having to resolve the 5 tuples of the original packets that are deeply encapsulated in the tunnel encapsulation.

The APN attribute allows to simplify the policy control at every policy enforcement point within the network. The APN attribute allows to reducing each matching entry of policy filter since it is only one field and hardware resources are saved. Since APN attribute is relatively stable, it introduces the possibilities of eliminating the "stale" policy filter entries. In most cases, the APN attribute

is centralized configured and distributed to all the policy enforcement points, which saves the policy filter configurations per node and simplifies the OM.

The structured APN attribute allows to express fine granular service requirements, e.g. MKT-user-group/app-group, RD-user-group/app-group, latency.

The structured APN attribute allows to match to the evolving fine granular differentiated network capabilities, e.g. SR policy with low latency and high reliability guaranteed.

In a tunnel across multiple domains of the same operator using the APN attribute in the outer header the operator can easily support multiple services not just a single one in a particular domain as illustrated in the usecase illustration section.

When there is no APN, to achieve the same, now the operator may have two options: 1. Add all the policy identifiers at the tunnel headend with various further encapsulations and enforce the policies based on them at the intermediate policy enforcement nodes along the tunnel, 2. Resolve the original 5 tuples being encapsulated inside the tunnel which will be very costly and sometimes impossible.

Moreover, the policy enforcement table in the intermediate policy enforcement nodes is significantly reduced. Because before operator needs to resolve the 5 tuple but now with APN, operator only needs to read the APN attribute in one field of the outer header.

Since the 5 tuples of the traffic are changing frequently due to service deployment or management issues the policy enforcement table in the policy enforcement nodes is not stable and there is always a lot of stale entries in the table. But now since the APN attribute is a mapping of the 5 tuples operator will have a relatively stable policy enforcement table on their nodes.

8. IANA Considerations

This document does not include an IANA request.

9. Security Considerations

In the APN work, in order to reduce the privacy and security issues, the following specifications are defined:

[S1]. The APN attribute MUST be conveyed along with the tunnel information in the APN domain. The APN attribute is encapsulated and removed at the APN-Edge.

[S2]. The APN ID (including the Application Group ID and the User Group ID) MUST be acquired from the existing available information in the packet header without interference into the payload.

According to the above specifications, the APN attribute is only produced and used locally within the APN domain without the involvement of the host/application side.

In order to prevent the malicious attack through the APN attribute, the following policies can be configured at the network devices of the APN domain:

[P1]. If the APN attribute is conveyed without the tunnel information, the packet MUST be dropped.

[P2]. If the APN attribute is not known to the APN domain, it should trigger the alarm information. The packet can be forwarded without being processed or dropped depending on the local policy.

[P3]. If the network service requirements exceed the specification for the specific Application Group ID and/or User Group ID, it should trigger the alarm information. The packet should be discarded to prevent abusing of the resources.

[P4]. There should be rate-limiting policy at the APN-Edge to prevent the traffic belonging to a specific Application Group ID and/or User Group ID from exceeding the preset limit.

10. Acknowledgements

The authors would like to acknowledge Robert Raszuk (Bloomberg LP), and Yukito Ueno (NTT Communications Corporation) for their valuable reviews and comments.

11. Co-authors

Kentaro Ebisawa
Toyota Motor Corporation
Japan

Email: ebisawa@toyota-tokyo.tech

Stefano Previdi
Huawei Technologies
Italy

Email: stefano@previdi.net

James N Guichard
Futurewei Technologies Ltd.
USA

Email: jguichar@futurewei.com

12. Contributors

Daniel Bernier
Bell Canada

Email: daniel.bernier@bell.ca

Chongfeng Xie
China Telecom

Email: xiechf@chinatelecom.cn

Feng Yang
China Mobile

Email: yangfeng@chinamobile.com

Zhuangzhuang Qin
China Unicom

Email: qinzhuangzhuang@chinaunicom.cn

Chang Liu
China Unicom

Email: liuc131@chinaunicom.cn

Gyan Mishra
Verizon

Email: hayabusagsm@gmail.com

Luis M. Contreras
Telefonica

Email: contreras.ietf@gmail.com

Luc-Fabrice Ndifor Ngwa
MTN

Email: Luc-Fabrice.Ndifor@mtn.com

13. References

13.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.
- [RFC8578] Grossman, E., Ed., "Deterministic Networking Use Cases", RFC 8578, DOI 10.17487/RFC8578, May 2019, <<https://www.rfc-editor.org/info/rfc8578>>.
- [RFC7665] Halpern, J., Ed. and C. Pignataro, Ed., "Service Function Chaining (SFC) Architecture", RFC 7665, DOI 10.17487/RFC7665, October 2015, <<https://www.rfc-editor.org/info/rfc7665>>.
- [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>>.
- [I-D.li-apn-problem-statement-usecases]
Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z., Mishra, G., Ebisawa, K., Previdi, S., and J. N. Guichard, "Problem Statement and Use Cases of Application-aware Networking (APN)", draft-li-apn-problem-statement-usecases-05 (work in progress), December 2021.
- [I-D.peng-apn-security-privacy-consideration]
Peng, S., Li, Z., Voyer, D., Li, C., Liu, P., and C. Cao, "APN Security and Privacy Considerations", draft-peng-apn-security-privacy-consideration-02 (work in progress), June 2021.

13.2. Informative References

- [RFC3272] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, DOI 10.17487/RFC3272, May 2002, <<https://www.rfc-editor.org/info/rfc3272>>.

Authors' Addresses

Zhenbin Li
Huawei Technologies
China

EMail: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
China

EMail: pengshuping@huawei.com

Daniel Voyer
Bell Canada
Canada

EMail: daniel.voyer@bell.ca

Cong Li
China Telecom
China

EMail: licong@chinatelecom.cn

Peng Liu
China Mobile
China

EMail: liupengyjy@chinamobile.com

Chang Cao
China Unicom
China

EMail: caocl5@chinaunicom.cn

Gyan Mishra
Verizon Inc.
USA

EMail: gyan.s.mishra@verizon.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 10 October 2022

Z. Li
S. Peng
Huawei Technologies
S. Zhang
China Unicom
8 April 2022

Application-aware Networking (APN) Header
draft-li-apn-header-02

Abstract

This document defines the application-aware networking (APN) header which can be used in a variety of data planes.

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 10 October 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	2
2. Requirements Language	2
3. Terminologies	3
4. Application-aware Networking Header	3
5. APN ID	7
6. APN Parameters	7
7. IANA Considerations	9
7.1. APN ID Types	10
7.2. APN Parameter Types	10
8. Acknowledgements	10
9. Security Considerations	11
10. References	11
10.1. Normative References	11
10.2. Informative References	11
Authors' Addresses	12

1. Introduction

Application-aware Networking (APN) is introduced in [I-D.li-apn-framework] and [I-D.li-apn-problem-statement-usecases]. APN conveys an attribute with data packets in the network and makes the network aware of fine-grain requirements at the user group and application group levels.

Such an attribute is acquired, constructed in a structured value, and then encapsulated in the packets. Such a structured value is treated as an opaque object in the network, to which the network operator applies policies in various nodes/service functions along the path and provides corresponding services.

This structured attribute can be encapsulated in various data planes adopted within a Network Operator's controlled and limited domain, e.g. MPLS, VXLAN, SR/SRv6 and other tunnel technologies, which waits to be further specified.

This document defines the application-aware networking (APN) header which can be used in different data planes. The typical data planes include the MPLS data plane and IPv6 data plane..

2. Requirements Language

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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminologies

APN: Application-aware Networking

APN Attribute: Application-aware Networking Attribute, including APN ID and APN Parameters. It can be added at the edge devices of an APN domain along with the tunnel encapsulation.

APN ID: Application-aware Networking ID, including Application Group ID and User Group ID.

APN Para: Application-aware Networking Parameters, e.g., network performance parameters.

4. Application-aware Networking Header

A common header is defined and can be used in different data planes. The common header carries the APN attribute that is composed of APN ID and APN parameters.

This document defines three types of APN ID:

- Type 1 APN ID: it is 32 bits.
- Type 2 APN ID: it is 64 bits.
- Type 3 APN ID: it is 128 bits.

According to the types of APN ID, three types of APN headers are defined.

Type 1 APN Header:

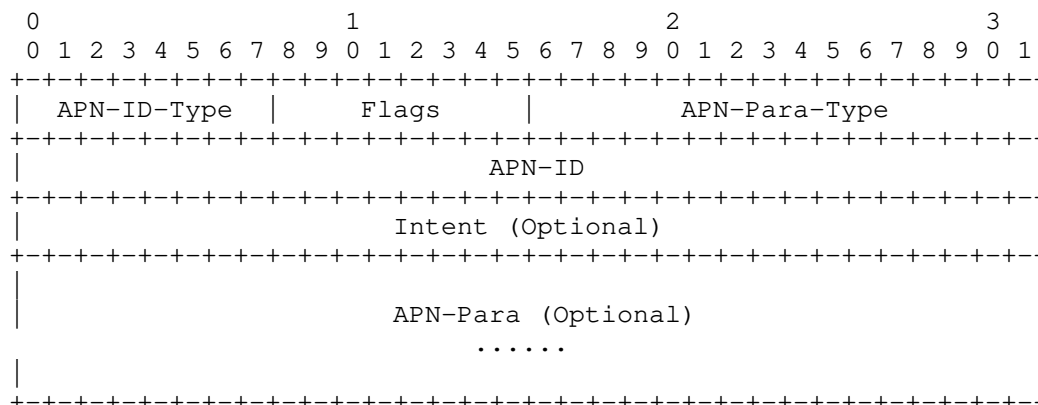


Figure 1. APN Header with Type 1 APN ID

In this type of APN Header, the length of the APN ID is 32 bits.

APN-ID-Type: An 8-bit identifier, indicates the type of APN ID.

Flags: An 8-bit field. The possible flags will be defined in the future versions of this document.

APN-Para-Type: A 16-bit map that specifies which APN parameters are specified for the APN ID. The APN-Para-Type value is a bitmap. The packing order of the APN parameters follows the bit order as specified in the APN-Para-Type bitmap field. The following bits are defined in this document, with details on each bit described in Section 6.

Bit 0 (Most significant bit) When set, indicates the presence of the bandwidth requirement.

Bit 1 When set, indicates the presence of the delay requirement.

Bit 2 When set, indicates the presence of the jitter requirement.

Bit 3 When set, indicates the presence of the packet loss rate requirement.

APN-ID: A 32-bit identifier.

Intent: A 32-bit identifier, represents a set of service requirements to the network.

APN-Para: A variable field including APN parameters. The presence of the APN parameters is indicated by the APN-Para-Type.

Type 2 APN Header

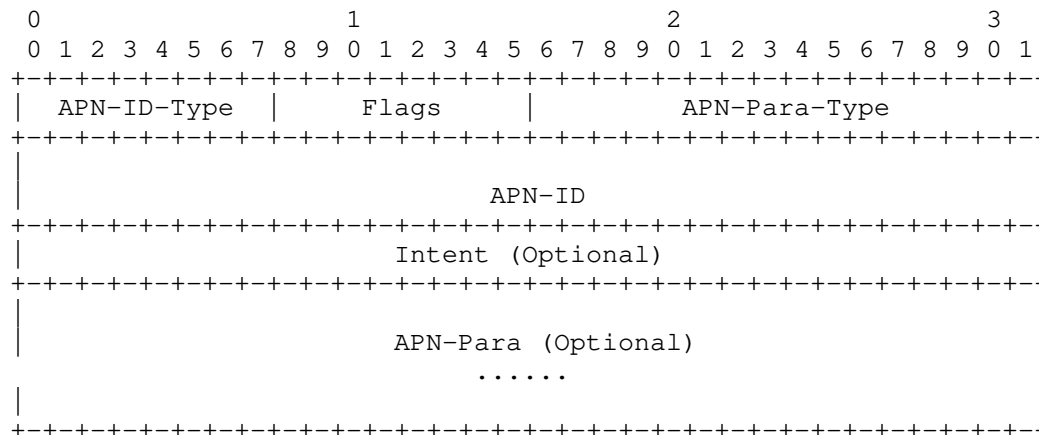


Figure 2. APN Header with Type 2 APN ID

In this type of APN Header, the length of the APN ID is 64 bits.

APN-ID-Type: An 8-bit identifier, indicates the type of APN ID.

Flags: An 8-bit field. The possible flags will be defined in the future versions of this document.

APN-Para-Type: A 16-bit map which specifies which APN parameters are specified for the APN ID. The following bits are defined in this document, with details on each bit described in the Section 6. The order of packing the data fields in each node data element follows the bit order of the APN-Para-Type field, as follows:

Bit 0 (Most significant bit) When set, indicates the presence of the bandwidth requirement.

Bit 1 When set, indicates the presence of the delay requirement.

Bit 2 When set, indicates the presence of the jitter requirement.

Bit 3 When set, indicates the presence of the packet loss rate requirement.

APN-ID: A 64-bit identifier.

Intent: A 32-bit identifier, represents a set of service requirements to the network.

APN-Para: A variable field including APN parameters. The presence of the APN parameters is indicated by the APN-Para-Type.

Type 3 APN Header

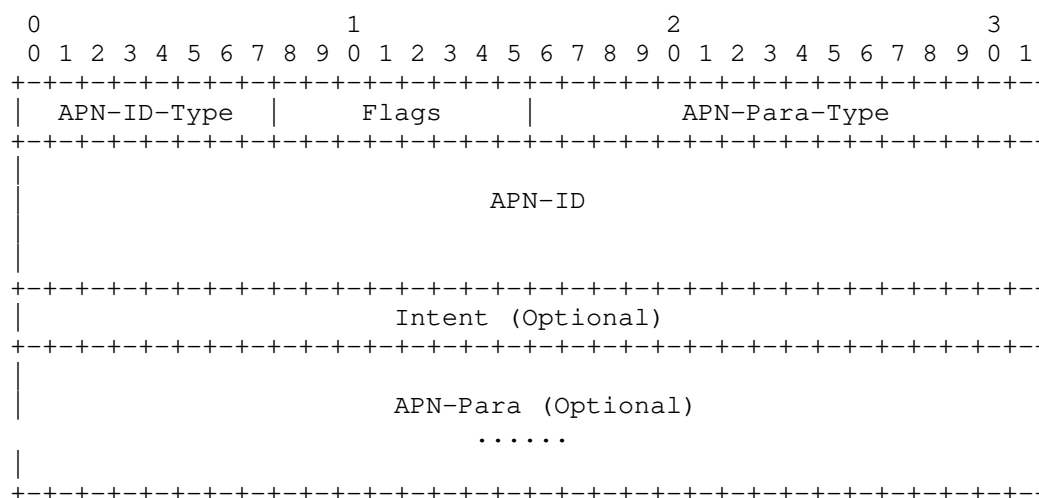


Figure 3. APN Header with Type 3 APN ID

In this type of APN Header, the length of the APN ID is 128 bits.

APN-ID-Type: An 8-bit identifier, indicates the type of APN ID.

Flags: An 8-bit field. The possible flags will be defined in the future versions of this document.

APN-Para-Type: A 16-bit map which specifies which APN parameters are specified for the APN ID. The following bits are defined in this document, with details on each bit described in the Section 6. The order of packing the data fields in each node data element follows the bit order of the APN-Para-Type field, as follows:

Bit 0 (Most significant bit) When set, indicates the presence of the bandwidth requirement.

Bit 1 When set, indicates the presence of the delay requirement.

Bit 2 When set, indicates the presence of the jitter requirement.

Bit 3 When set, indicates the presence of the packet loss rate requirement.

APN-ID: A 128-bit identifier.

Intent: A 32-bit identifier, represents a set of service requirements to the network.

APN-Para: A variable field including APN parameters. The presence of the APN parameters is indicated by the APN-Para-Type.

5. APN ID

The APN ID can be divided into three parts:

APP-Group-ID: Application Group ID

USER-Group-ID: User Group ID

Reserved: The reserved field.

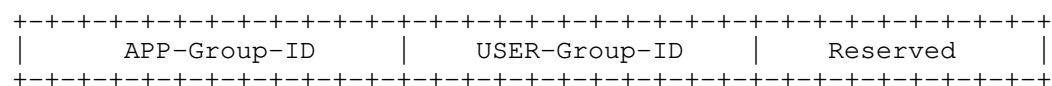


Figure 4. Structure of APN-ID

The lengths of the APP-Group-ID and the USER-Group-ID are variable. Their lengths must be configured and consistent within a specific APN domain.

6. APN Parameters

In the APN Header, the APN-Para-Type is a bit field to indicate the presence of corresponding APN parameters. When the bit is set, the corresponding APN parameter MUST exist in the APN Header. The length of each APN parameter is 32 bits. Thus it is easy to skip over unknown requirements.

Typical APN parameters are the parameters related with the network performance requirements as follows:

1. Bandwidth Requirement

This Bandwidth Requirement parameter indicates the minimum acceptable bandwidth for the APN traffic. The format of this parameter is shown in the following diagram:

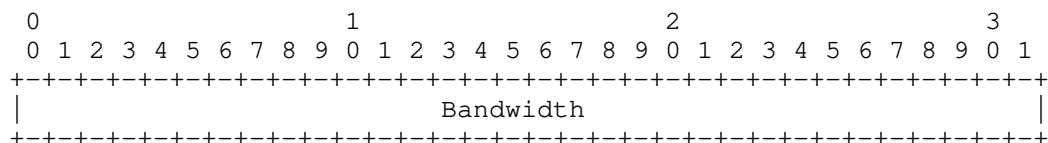


Figure 5. Bandwidth Requirement Parameter

where:

Bandwidth: This 32-bit unsigned integer field carries the bandwidth requirement in Mbps along the path.

2. Delay Requirement

This Delay Requirement parameter indicates the maximum acceptable delay. The format of this parameter is shown in the following diagram:

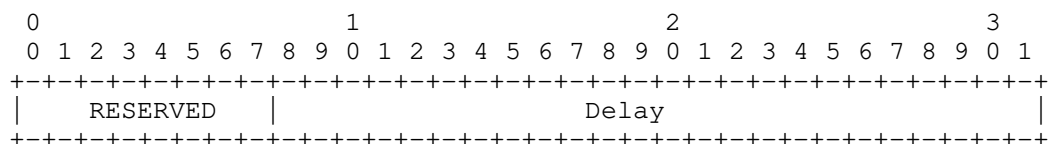


Figure 6. Delay Requirement Parameter

where:

RESERVED: This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Delay: This 24-bit field carries the delay requirements in microseconds, encoded as an unsigned integer value. When set to the maximum value 16,777,215 (16.777215 sec), then the delay is not constrained. This value is the highest delay that can be tolerated.

3. Delay Variation Requirement

This Delay Variation Requirement parameter indicates the maximum acceptable delay variation. The format of this parameter is shown in the following diagram:

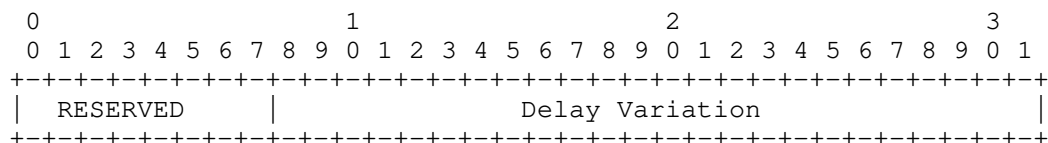


Figure 7. Delay Variation Parameter

where:

RESERVED: This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Delay Variation: This 24-bit field carries the delay variation requirements in microseconds, encoded as an unsigned integer value.

4. Packet Loss Rate Requirement

This Packet Loss Rate Requirement parameter indicates the maximum acceptable packet loss rate. The format of this parameter is shown in the following diagram:

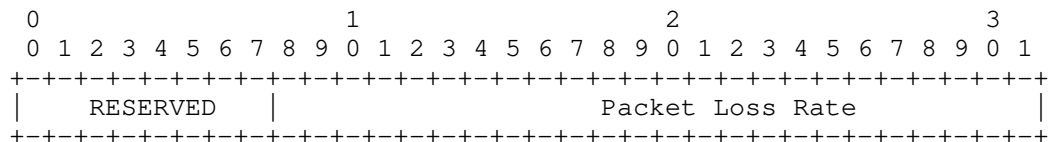


Figure 8. Packet Loss Rate Sub-TLV

where:

RESERVED: This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Packet Loss Rate: This 24-bit field carries packet loss rate requirement in packets per second as an unsigned integer. This value is the highest packet-loss rate that can be tolerated.

7. IANA Considerations

These IANA Considerations conform to [RFC8126].

IANA is requested to create the following new registries on a new "Application-Aware Networking (APN)" webpage.

7.1. APN ID Types

IANA is requested to create the following registry on the Application-Aware Networking (APN) Attribute webpage:

Name: APN ID Types

Registration Procedure: IETF Review

Reference: [this document]

Value	Description	Reference
0	reserved	
1	Type 1 APN ID	[this document]
2	Type 2 APN ID	[this document]
3	Type 3 APN ID	[this document]
4-254	unassigned	
255	reserved	

7.2. APN Parameter Types

IANA is requested to create the following registry on the Application-Aware Networking (APN) Attribute webpage:

Name: APN Parameter Types

Registration Procedure: IETF Review

Reference: [this document]

Bit	Description	Reference
0	Bandwidth requirement	[this document]
1	Delay requirement	[this document]
2	Jitter requirement	[this document]
3	Packet loss requirement	[this document]
4-15	unassigned	

8. Acknowledgements

The suggestions of the following are gratefully acknowledged: Stefano Previdi, Adrian Farrel, Donald Eastlake.

9. Security Considerations

The Security Considerations described in [I-D.li-apn-problem-statement-usecases] and [I-D.peng-apn-security-privacy-consideration] can be referred to.

10. References

10.1. Normative References

- [I-D.li-apn-framework]
Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C., and G. Mishra, "Application-aware Networking (APN) Framework", Work in Progress, Internet-Draft, draft-li-apn-framework-05, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-framework-05.txt>>.
- [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>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [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>>.

10.2. Informative References

- [I-D.li-apn-problem-statement-usecases]
Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z., and G. Mishra, "Problem Statement and Use Cases of Application-aware Networking (APN)", Work in Progress, Internet-Draft, draft-li-apn-problem-statement-usecases-06, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-problem-statement-usecases-06.txt>>.
- [I-D.peng-apn-security-privacy-consideration]
Peng, S., Li, Z., Voyer, D., Li, C., Liu, P., and C. Cao, "APN Security and Privacy Considerations", Work in Progress, Internet-Draft, draft-peng-apn-security-privacy-consideration-02, 16 June 2021, <<https://www.ietf.org/archive/id/draft-peng-apn-security-privacy-consideration-02.txt>>.

[RFC3272] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, DOI 10.17487/RFC3272, May 2002, <<https://www.rfc-editor.org/info/rfc3272>>.

Authors' Addresses

Zhenbin Li
Huawei Technologies
Beijing
100095
China
Email: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Beijing
100095
China
Email: pengshuping@huawei.com

Shuai Zhang
China Unicom
Beijing
China
Email: zhangs366@chinaunicom.cn

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 10 October 2022

Z. Li
S. Peng
Huawei Technologies
C. Xie
China Telecom
8 April 2022

Application-aware IPv6 Networking (APN6) Encapsulation
draft-li-apn-ipv6-encap-04

Abstract

Application-aware IPv6 Networking (APN6) makes use of IPv6 encapsulation to convey the APN Attribute along with data packets and make the network aware of data flow requirements at different granularity levels. The APN attribute can be encapsulated in the APN header. This document defines the encapsulation of the APN header in the IPv6 data plane.

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 10 October 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	2
2. Requirements Language	3
3. Terminologies	3
4. The APN Option	3
5. Locations for the APN Option	4
5.1. IPv6 Hop-by-Hop Options Header (HBH)	4
5.2. IPv6 Destination Options Header (DOH)	4
6. APN TLV for the SRH	4
7. IANA Considerations	5
7.1. IPv6 Header Option	5
7.2. SRH TLV Type	6
8. Security Considerations	6
9. References	6
9.1. Normative References	6
9.2. Informative References	7
Authors' Addresses	7

1. Introduction

Application-aware Networking (APN) is introduced in [I-D.li-apn-framework] and [I-D.li-apn-problem-statement-usecases]. APN conveys an attribute along with data packets into the network and make the network aware of data flow requirements at different granularity levels. Such an attribute is acquired, constructed as a structured value, and then encapsulated in the packets. Such a structured value is treated as an opaque object in the network, to which the network operator applies policies in various nodes/service functions along the path, providing corresponding services.

[I-D.li-apn-header] defines the application-aware networking (APN) header which can be used in different data planes to carry the APN attribute. This document defines the encapsulation of the APN header in the IPv6 data plane.

2. Requirements Language

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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminologies

APN: Application-aware Networking

APN6: Application-aware IPv6 Networking, i.e., the data plane of APN is IPv6

APN Attribute: Application-aware information. It is added at the edge devices of an APN domain along with any tunnel encapsulation.

APN ID: Application-aware Networking ID

APN Para: Application-aware Networking Parameters

SRH: Segment Routing Header RFC 8754 [RFC8754]

4. The APN Option

To support Application-aware IPv6 networking, one IPv6 Header option RFC 8200 [RFC8200], the APN option, is defined.

The APN option has the following format:

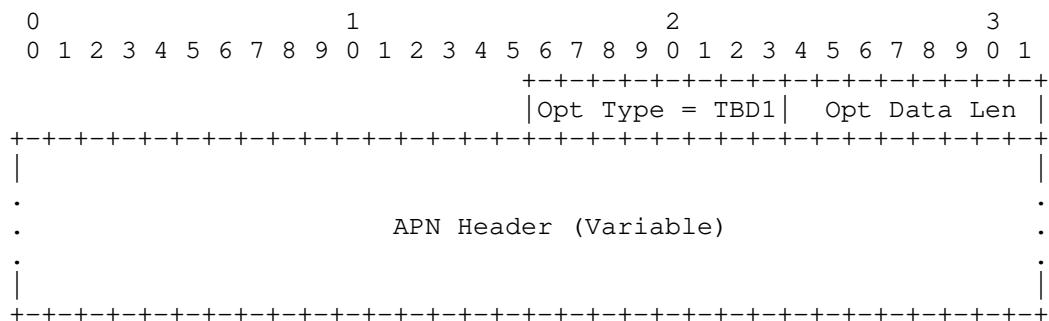


Figure 1. The APN Option

where:

- o Opt Type: Type value is TBD1, an 8-bit unsigned integer. Identifier of the type of this APN Option.
- o Opt Data Len: An 8-bit unsigned integer. Length of the Option Data field of this option, that is, length of the APN header.
- o APN Header: Option-Type-specific data. It carries the APN header. Variable-length field as specified in [I-D.li-apn-header].

5. Locations for the APN Option

The APN IPv6 Header option can be placed in two locations in an IPv6 packet header RFC 8200 [RFC8200] depend upon the scenario and implementation requirements. These are defined in the subsections below.

5.1. IPv6 Hop-by-Hop Options Header (HBH)

The APN option can be carried in the IPv6 Hop-by-Hop Options Header. By using the HBH Options Header, the information carried can be read by every node along the path.

5.2. IPv6 Destination Options Header (DOH)

The APN option can be carried in the IPv6 Destination Options Header. By using the DOH Options Header, the information carried can be read by the destination node but would not normally be seen by other nodes along the path.

6. APN TLV for the SRH

[RFC8754] defines the segment routing header (SRH) and the SRH TLV. The SRH TLV provides meta-data for segment processing. The APN header can be placed in the SRH as the value of one type of SRH TLV following the Segment List. By using the SRH, the information carried can be read by the specified segment destinations along the SRv6 path.

The APN TLV is OPTIONAL and has the following format:

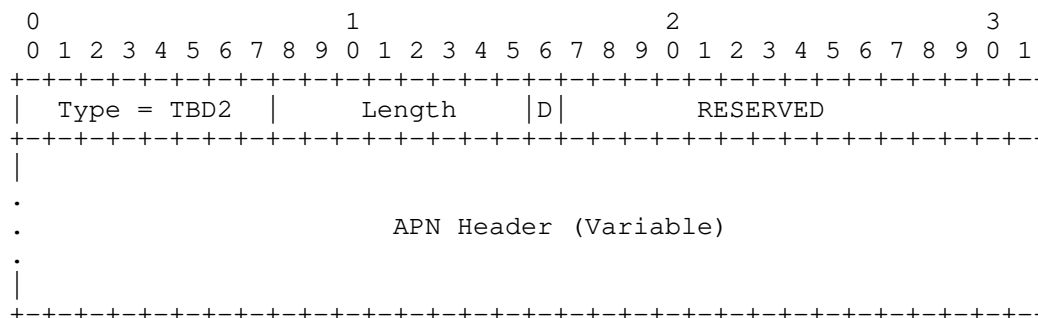


Figure 2. The APN SRH TLV

where:

- o Type: TBD2 (suggested value 5).
- o Length: The length of the variable length data in bytes.
- o D: 1 bit. When it is set, it indicates the Destination Address verification is disabled due to use of a reduced segment list.
- o RESERVED: 15 bits. MUST be 0 on transmission and ignored on receipt.
- o APN Header: It carries the APN header as specified in [I-D.li-apn-header]. A variable-length field.

7. IANA Considerations

IANA is requested to assign two code points as below.

7.1. IPv6 Header Option

IANA is requested to assign an IPv6 Header Option as follows:

Hex Value	Binary Value			Description	Reference
	act	chg	rest		
TBD1	00	0	xxxxxx	Application-aware Networking	[this document]

RFC Editor / IANA Note: To be removed when RFC is published. xxxxx is the binary for the bottom 5 bits of TBD1.

7.2. SRH TLV Type

IANA is requested to assign an SRH TLV Type from the range of type values for TLVs that do not change en route (2-127) as follows:

Value	Description	Reference
TBD2	Application-aware Networking	[this document]

8. Security Considerations

The Security Considerations are described in [I-D.li-apn-problem-statement-usecases].

9. References

9.1. Normative References

- [I-D.li-apn-framework]
Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C., and G. Mishra, "Application-aware Networking (APN) Framework", Work in Progress, Internet-Draft, draft-li-apn-framework-05, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-framework-05.txt>>.
- [I-D.li-apn-header]
Li, Z., Peng, S., and S. Zhang, "Application-aware Networking (APN) Header", Work in Progress, Internet-Draft, draft-li-apn-header-01, 6 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-header-01.txt>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.

9.2. Informative References

[I-D.li-apn-problem-statement-usecases]
Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z., and G. Mishra, "Problem Statement and Use Cases of Application-aware Networking (APN)", Work in Progress, Internet-Draft, draft-li-apn-problem-statement-usecases-06, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-problem-statement-usecases-06.txt>>.

Authors' Addresses

Zhenbin Li
Huawei Technologies
Beijing
100095
China
Email: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Beijing
100095
China
Email: pengshuping@huawei.com

Chongfeng Xie
China Telecom
China
Email: xiechf@chinatelecom.cn

RTGWG Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2022

S. Liu
China Mobile
H. Zheng
Huawei Technologies
October 25, 2021

Accessing Cloud via Optical Network Problem Statement
draft-liu-rtgwg-optical2cloud-problem-statement-00

Abstract

This document describes the scenarios and requirements for the Cloud accessing through optical network, as a complementary functionality of the network and cloud coordination. The problem from optical perspective is different with packet, and statement is made in this document.

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 April 28, 2022.

Copyright Notice

Copyright (c) 2021 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 Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Requirements Language	3
2. Scenarios	3
2.1. Multi-cloud accessing	3
2.2. High-quality leased line	4
2.3. Cloud virtual reality	5
3. Requirement and Problem statement	5
3.1. LxVPN of optical networks for multiple-to-multiple access .	5
3.2. Small Granularity Switching	6
3.3. High-performance and high-reliability	6
4. Manageability Considerations	6
5. Security Considerations	7
6. IANA Considerations	7
7. References	7
7.1. Normative References	7
7.2. Informational References	7
Authors' Addresses	7

1. Introduction

The cloud-related applications is becoming popular and wider deployed, in enterprises and vertical industries. Companies with multi-campus are interconnected together with the remote cloud, for the purpose of storage and computation. Such cloud services require high-level experiences including high availability, low latency, on-demand adjustment and so on.

Optical is playing an important role in the transport network, with its own large bandwidth and low latency feature. Based on the TDM switching technology, the data transportation in optical networks does not have any queuing problem to solve and can perfectly avoid congestion. Such features can drastically improve the users experience on the service quality.

Optical network is considered as the transportation solution for long-distance. This feature is also suitable for the cloud interconnections, especially when there is demand for large bandwidth.

[I-D.ietf-rtgwg-net2cloud-problem-statement] and [I-D.ietf-rtgwg-net2cloud-gap-analysis] gave a detailed description on the coordination requirements between the network and the cloud, and

it is expected the description in this document can be used as a complementary from the optical perspective.

1.1. Requirements Language

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.

2. Scenarios

With the prevalence of cloud services, enterprises services, home services such as AR/VR, accessing clouds with optical networks is increasingly attractive and becoming an option for the users. Following scenarios provide a few typical applications.

2.1. Multi-cloud accessing

Cloud services are usually supported by multiple interconnected data centers (DCs). Besides the on-demand, scalable, high available and uses-based billing, mentioned in [I-D.ietf-rtgwg-net2cloud-problem-statement], there are also needs for Data Centre Interconnect (DCI) about high requirements on capacity, latency, and flexible scheduling. This use case requires specific capabilities of advanced OTN (Optical Transport Network) for DCIs.

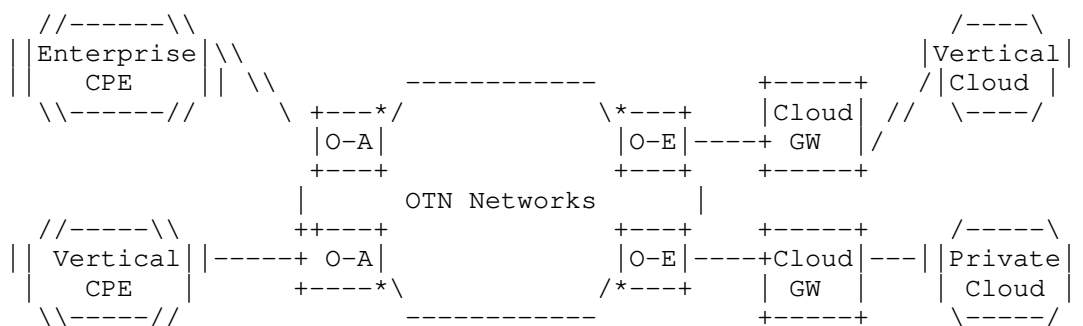


Figure 1: Cloud Accessing through Optical Network

A data center is a physical facility consisting of multiple bays of interconnected servers, that performs computing, storage, and communication needed for cloud services. Infrastructure-as-a-service

may be deployed in both public and private clouds, where virtual servers and other virtual resources are made available to users on demand and by self-service.

One typical scenario is the intra-city DCs, which communicate with each other via the intra-city DCI network to meet the high availability requirements. The active-active and Virtual Machine (VM) migration services which require low latency are provided by the intra-city DCI network. The intra-city DCI network supports the public and/or the private cloud services, such as video, games, desktop cloud, and cloud Internet cafe services. To ensure low latency, intra-city DCI network is deployed in the same city or adjacent cities. The distance is typically less than 100 km and more likely less than 50km. One city may have several large DCs.

DCs are ideally interconnected through Layer 2/3 switches or routers with full mesh connectivity. However, to improve interaction efficiency as well as service experience, OTN is also evaluated as an option to be used for DC interconnection.

There are three kinds of the connection relationship, point to point access, single to multiple point access, and multiple to multiple point access. Different types of connections are referring different shapes, single point accessing single cloud, single point accessing multiple clouds and multiple points accessing multiple clouds.

2.2. High-quality leased line

The high quality private line provides high security and reliability and is suitable to ensure the end-to-end user experience for large enterprises such as financial, medical centers and education customers. The main advantages and drivers of the high quality private line are as follows.

- o High quality private lines provide large bandwidth, low latency, secure and reliable for any type of connection.
- o Accelerate the deployment of cloud services. The high-quality and high-security of the private line connecting to the cloud can enable enterprises to move more core assets to the cloud and use low-latency services on the cloud. Cloud-based deployment helps enterprises reduce heavy asset allocation and improve energy saving, so that enterprises can focus on their major business.
- o Reduce operator's CAPEX and OPEX. The end-to-end service provisioning system enables quick provisioning of private line services and improves user experience. Fault management can be done from the device level to reduce the complexity of location.

- o Enable operators to develop value-added services by providing enterprise users with latency maps, availability maps, comprehensive SLA reports, customized latency levels, and dynamic bandwidth adjustment packages.

2.3. Cloud virtual reality

Cloud Virtual Reality (VR) offloads computing and cloud rendering in VR services from local dedicated hardware to a shared cloud infrastructure. Cloud rendered video and audio outputs are encoded, compressed, and transmitted to user terminals through fast and stable networks. In contrast to current VR services, where good user experience primarily relies on the end user purchasing expensive high-end PCs for local rendering, cloud VR promotes the popularization of VR services by allowing users to enjoy various VR services where rendering is carried out in the cloud.

Cloud VR service experience is impacted by several factors that influence the achieved sense of reality, interaction, and immersion, which are related to the network properties, e.g. bandwidth, latency and packet loss. The network performance indicators, such as bandwidth, latency, and packet loss rate, need to meet the requirements to realize a pleasurable experience.

The current network may be able to support early versions of cloud VR (e.g. 4K VR) with limited user experience, but will not meet the requirements for large scale deployment of cloud VR with enhanced experience (e.g. Interactive VR applications, cloud games). To support more applications and ensure a high-quality experience, much higher available and guaranteed bandwidth (e.g. larger than 1 Gbps), lower latency (e.g. less than 10 ms) and lower jitter (e.g. less than 5 ms) are required.

3. Requirement and Problem statement

3.1. LxVPN of optical networks for multiple-to-multiple access

To establish MP2MP connections, TDM transport technologies, like OTN, are adopting packet features. Some OTN equipments have adopted packet processing functions, such as packet switching, MPLS VPN, etc., which could provide an underlay performance guaranteed TDM channel for cloud accessing, as an alternative of packet-based connections.

3.2. Small Granularity Switching

According to the ITU-T G.709 recommendation, the OTN is providing TDM based connection with a granularity 1.25Gbps, which is more than the demand for normal user. Most of the leased line is requesting a bandwidth less than 10Mbps, and the request from big enterprises are usually on the level of 100Mbps. Therefore, most of the leased lines are with small granularity in the field.

The SDH was a good complementary of OTN for small granularity solution, but SDH devices are gradually removed from the network due to End of Services. As SDH networks gradually phase out, service providers start to think about how to utilize OTN networks to transmit small-granularity high-value SDH services. The OSU (optical service unit) is proposed to solve the problem.

At ITU-T, two work items, G.sub1G.sup and G.OSU, have been initiated aiming to enable OTN to support small-granularity services of 2M-1Gb/s. For G.OSU, the general idea is to put small granularity services into OSU containers, and then put OSU containers into OPU payload areas. OSU containers are flagged by Tributary Port Number (TPN) tags located at the overhead of the OSU containers. At the intermediate nodes, OSUs can be switched to different directions based on the TPN tags in the overhead. Given the development of OSU, the OTN is expected to be able to carry small granularity service and create end-to-end optical connections.

3.3. High-performance and high-reliability

To support the above-mentioned applications some of the network properties are critical to promise the Quality of Services (QoS). For instance, high bandwidth (e.g. larger than 1 Gbps), low latency (e.g. no more than 10 ms) and low jitter (e.g. no more than 5 ms), are required for Cloud VR. In addition, small-granularity container is required to improve the efficiency of the networks.

It is also critical to support highly reliable DCI for cloud services. With advanced optical transport network protection and automatic recovery technologies, services can still run properly even fiber cuts occur in the DCI network. Specific protection and restoration schemes are required, to provide high reliability for the networks.

4. Manageability Considerations

TBD.

5. Security Considerations

TBD.

6. IANA Considerations

This document requires no IANA actions.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

7.2. Informational References

- [I-D.ietf-rtgwg-net2cloud-gap-analysis]
Dunbar, L., Malis, A. G., and C. Jacquenet, "Networks Connecting to Hybrid Cloud DCs: Gap Analysis", draft-ietf-rtgwg-net2cloud-gap-analysis-07 (work in progress), July 2020.
- [I-D.ietf-rtgwg-net2cloud-problem-statement]
Dunbar, L., Consulting, M., Jacquenet, C., and M. Toy, "Dynamic Networks to Hybrid Cloud DCs Problem Statement", draft-ietf-rtgwg-net2cloud-problem-statement-11 (work in progress), July 2020.

Authors' Addresses

Sheng Liu
China Mobile
China

Email: liushengwl@chinamobile.com

Haomian Zheng
Huawei Technologies
H1, Xiliu Beipo Village, Songshan Lake,
Dongguan, Guangdong 523808
China

Email: zhenghaomian@huawei.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 31 October 2022

S. Peng
Z. Li
S. Fang
Huawei Technologies
Y. Cui
Tsinghua University
29 April 2022

Dissemination of BGP Flow Specification Rules for APN
draft-peng-apn-bgp-flowspec-01

Abstract

A BGP Flow Specification is an n-tuple consisting of several matching criteria that can be applied to IP traffic. Application-aware Networking (APN) is a framework, where APN data packets convey APN attribute including APN ID and/or APN Parameters. The dynamic Flow Spec mechanism for APN is designed for the new applications of traffic filtering in an APN domain as well as the traffic control and actions at the policy enforcement points in this domain. These applications require coordination among the ASes within a service provider.

This document specifies a new BGP Flow Spec Component Type in order to support APN traffic filtering. The match field is the APN ID. It also specifies traffic filtering actions to enable the creation of the APN ID in the outer tunnel encapsulation when matched to the corresponding Flow Spec rules.

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 31 October 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	3
2. Requirements Language	4
3. Terminologies	4
4. Flow Specifications for APN	4
5. Component Type for APN	5
5.1. APN ID - Type TBD1	5
5.2. Encoding Example	5
6. Traffic Filtering	6
6.1. Ordering of Flow Specifications	6
6.2. Encoding format of the Grouping Identifier Extend Community Sub-Type TBD2	7
6.3. Usage Principles	7
6.4. Usage example	8
7. Traffic Filtering Actions	10
7.1. Traffic Marking (traffic-marking-apn) Sub-Type TBD3 . . .	11
7.2. Traffic Marking (traffic-marking-apn-partial) Sub-Type TBD4	12
7.3. Inherit (inherit-apn) Sub-Type TBD5	13
7.4. Stitch (stitch-apn) Sub-Type TBD6	13
8. IANA Considerations	14
8.1. Flow Spec Component - APN ID	14
8.2. Opaque Extended Community - Grouping Identifier	15
8.3. Extended Community Flow Specification Actions	15
9. Acknowledgements	15
10. Security Considerations	16
11. References	16
11.1. Normative References	16
11.2. Informative References	17
Authors' Addresses	18

1. Introduction

A Flow Specification (Flow Spec) is an n-tuple consisting of several matching criteria that can be applied to IP traffic [RFC8955]. The Flow Spec conveys match conditions (each may include several components) which are encoded using MP_REACH_NLRI and MP_UNREACH_NLRI attributes [RFC4760], while the associated actions such as redirect and traffic marking are encoded in BGP Extended Communities [RFC4360][RFC5701]. The IPv4 NLRI component types and traffic filtering actions sub-types are described in [RFC8955], while the IPv6 related are described in [RFC8956]. [I-D.ietf-idr-flowspec-l2vpn] extends the flow-spec rules and actions for Ethernet Layer 2 and L2VPN. The corresponding (AFI, SAFI) pairs are defined by IANA, respectively. [I-D.hares-idr-flowspec-v2] specifies BGP Flow Specification Version 2.

Application-aware Networking (APN) is introduced in [I-D.li-apn-framework] and [I-D.li-apn-problem-statement-usecases]. APN data packets convey the APN attribute (incl. APN ID and/or APN Parameters). The APN ID is a structured value, treated as an opaque object in the network, to which the network operator applies policies in various nodes/service functions along the path so to provide corresponding services. For an IPv6 network, a design proposal of such structured value is provided by [I-D.li-apn-header][I-D.li-apn-ipv6-encap]. The APN attribute can be encapsulated in various data planes adopted within a Network Operator controlled limited domain, e.g. IPv6, MPLS, and other tunnel technologies, which wait to be further specified.

With APN, it becomes possible to apply various policies in different nodes along a network path onto a traffic flow overall in a more efficient way, that is, at the headend to steer into corresponding path, at the midpoint to collect corresponding performance measurement data, and at the service function to execute particular policies. Prior to APN, there was no efficient way to realize this composite network service provisioning along the path.

This document specifies a new BGP Flow Spec Component Type to support APN traffic filtering. The match field is the APN APN ID [I-D.li-apn-framework]. It also specifies traffic filtering actions to enable the creation of the APN ID in the outer tunnel encapsulation when matched to the corresponding Flow Spec rules.

Depends upon specific deployment requirements, the functions specified in this draft can also be applied on BGP Flow Specification Version 2, which will be specified in the future versions.

2. Requirements Language

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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminologies

APN: Application-aware Networking

APN ID: APN Identifier

AS: Autonomous System

Flow Spec: Flow Specification

BGP-FS: Border Gateway Protocol (BGP) Flow Specification (FS)

4. Flow Specifications for APN

The APN framework is introduced in [I-D.li-apn-framework]. The Flow Spec for APN is shown in Figure 1, that is, the Controller is used to set up BGP connection with the policy enforcement points in an APN domain.

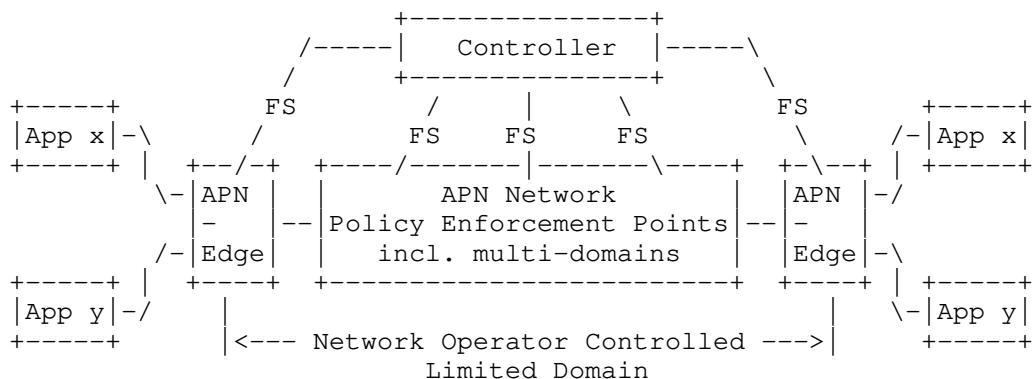


Figure 1. Flow Spec for APN

5. Component Type for APN

The IPv4 NLRI component types are defined in [RFC8955], while the IPv6 related are specified in [RFC8956]. This document defines a new component type for APN.

5.1. APN ID - Type TBD1

Encoding: <type (1 octet), length (1 octet), mask (variable), APN ID (variable)>

Defines the APN ID to match. The mask is used to indicate the bits of the APN ID carried in the packet which are used to match against the APN ID value in this Flow Spec component.

type (1 octet): This indicates the new component type TBD1.

length (1 octet): This indicates the length of the mask and the length of the APN ID. The mask and the APN ID have the same length.

mask (variable): This indicates the bits of the APN ID carried in the data packet which are used to match.

APN ID (variable): This indicates the APN ID that is used for the match.

5.2. Encoding Example

Since the APN ID is a structured value, the mask in the Flow Spec is used to enable flexible matching of the particular parts of the APN ID.

As an example, shown in Figure 2, the APN ID in the data packet contains two parts, the APP Group ID (0x300A) and User Group ID (0x0C08). In the Flow Spec, the mask is 0xFFFF0000 and the APN ID is 0x300A0000. Processing the match of the APN ID component is done by using the mask (0xFFFF0000) to indicate the bits of the APN ID carried in the packet to be matched against the one carried in the Flow Spec (0x300A0000). The result of this example is a successful match.

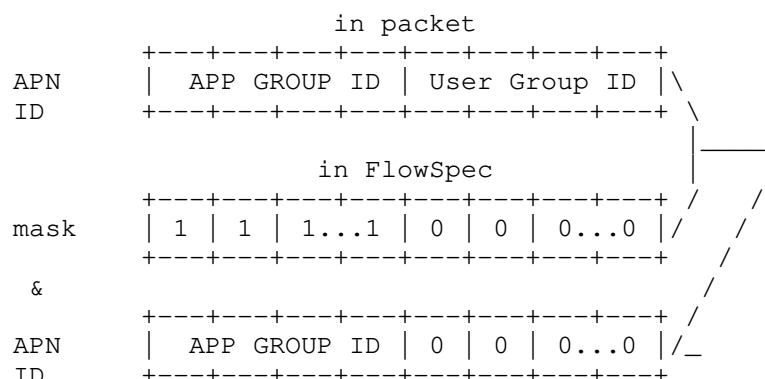


Figure 2. The match example of the APN ID component

6. Traffic Filtering

Traffic filtering policies have been traditionally considered to be relatively static. The dynamic Flow Spec mechanism for APN is designed for the new applications of traffic filtering in an APN domain as well as the traffic control and actions on the policy enforcement points in this domain. These applications require coordination among the ASes within a service provider. The new component and encoding are defined in Section 4. The actions are defined in this section.

6.1. Ordering of Flow Specifications

More than one Flow Specification rule may match a particular traffic flow at a node. The co-existing rules are mixed and need to be effectively organized. However, there is still no efficient way to achieve such classification. Thus, it is necessary to specify the grouping mechanism for the Flow Specification rules to be matched in a desired order as well as the actions being applied to a particular traffic flow. This ordering function is such that it does not depend on the arrival order of the Flow Specification via BGP and thus is consistent in the network [RFC8955].

The definition of this ordering is very important to the Flow Spec for APN because of the following reasons.

1. There can be other co-existing Flow Spec rules (e.g. based on 5-tuple) rather than only APN Flow Spec rules (i.e. based on APN ID).
2. The different parts of the APN ID can be determined by the different Flow Spec rules.

Therefore, the ordering of the Flow Spec rules for APN needs to be clearly specified.

6.2. Encoding format of the Grouping Identifier Extend Community Sub-Type TBD2

We define a Grouping Identifier Opaque Extend Community [RFC4360] (Sub-Type = TBD2) carrying both Group ID (2 octets) and Sub-group ID (2 octets) and indicating the grouping of the Flow Spec rules it accompanies.

The encoding format of the Grouping Identifier Opaque Extend Community is as follows.

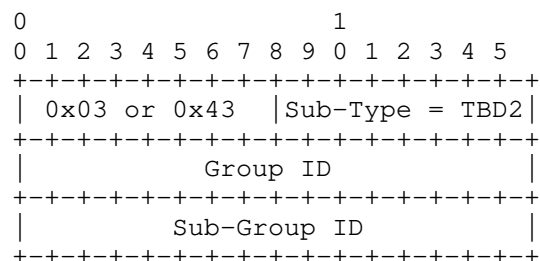


Figure 3: Encoding of the Grouping Identifier Extend Community

6.3. Usage Principles

The following principles are defined.

1. Within a sub-group, the order is the same as the previously defined.
- * If the traffic-action Extended Community is carried and the Terminal Action (T, bit 47) [RFC8955] is not set, when one condition in this sub-group is matched, the evaluation of any subsequent flow specifications within this sub-group stops; if T is set, then the evaluation continues;
- * If the traffic-action Extended Community is not carried, when one condition in this sub-group is matched, the evaluation of any subsequent flow specifications within this sub-group stops;

2. Between sub-groups, the sub-group is ordered by increasing Sub-group ID, when the evaluation in one sub-group stops or finishes, it will start the evaluation in the following sub-group if there are any sub-groups.

3. Between groups, the group is ordered by increasing Group ID, if at least one condition in this group is matched, when the evaluation of the flow specifications within the group reaches the end, the evaluation stops so the evaluation of the following group(s) will not start.

6.4. Usage example

At the APN Edge where the APN ID is created based on the Flow Specifications and encapsulated in the outer tunnel header [I-D.li-apn-framework], more than one Flow Specification rule condition may match a particular traffic flow. The different parts of the APN ID can be determined by the different Flow Spec rules. For example, as shown in Figure 4, the App Group ID is created by matching the 5-tuple components (e.g. destination IP address and transport layer ports), the User Group ID is created by matching the access ports, and the Reserved (R.) Group ID is created by matching the 5-tuple components.

Moreover, there are also other co-existing Flow Spec rules mixed at the node rather than only APN Flow Spec rules (i.e. based on APN ID). All the rules need to be effectively organized and applied to the particular traffic flow in a desired order.

In Figure 4, the Flow Specification rules for APN and other existing rules are categorized into two groups, and given Group ID = 1 and 2, respectively. The Flow Specification rules for creating different parts of the APN ID are categorized into three sub-groups, and given Sub-Group ID = 1, 2, and 3, respectively.

Based on the usage principles described in the above section, for the case of APN as shown in Figure 4, the usage principles are as follows,

1. Within a sub-group, the order is the same as the previously defined.

- * If the traffic-action Extended Community is carried and the Terminal Action (T, bit 47) [RFC8955] is not set, when one condition in this sub-group is matched, the evaluation of any subsequent flow specifications within this sub-group stops and the App Group ID is created; if T is set, then the evaluation continues and the App Group ID will be created if there is a match within this sub-group;
- * If the traffic-action Extended Community is not carried, when one condition in this sub-group is matched, the evaluation of any subsequent flow specifications within this sub-group stops and the App Group ID is created;

2. Between sub-groups, the sub-group is ordered with Sub-group ID, when the evaluation in the Sub-group ID = 1 stops or finishes, it will start the evaluation in the following Sub-group ID = 2 and create the User Group ID if matched, and then the Sub-group ID = 3 to create the R. Group ID if matched.

3. Between groups, the group is ordered with Group ID, if at least one condition in this Group ID = 1 is matched, when the evaluation of the flow specifications within the group reaches the end, the evaluation stops and the APN ID is created. The evaluation of the following group(s) will not start, that is, the Group ID = 0 will not be evaluated.

Group ID = 1, Sub-Group ID = 1

Rule (5-tuple)	App Group ID
Rule (5-tuple)	App Group ID
...	...
Rule (5-tuple)	App Group ID

Group ID = 1, Sub-Group ID = 2

Rule (ports)	User Group ID
Rule (ports)	User Group ID
...	...
Rule (ports)	User Group ID

+--+	
---	--

Figure 4: Usage of Grouping Identifier Extended Community for APN

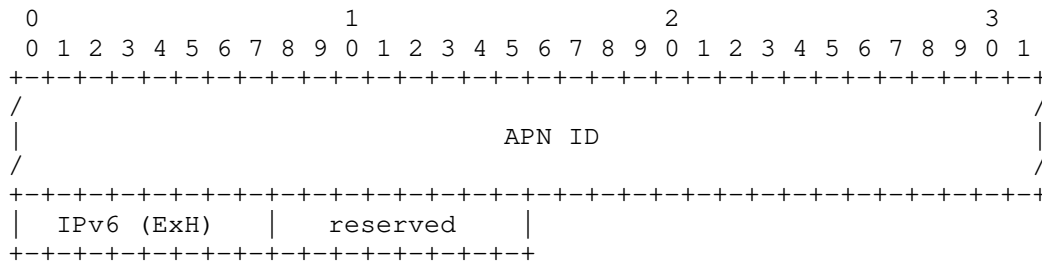
7. Traffic Filtering Actions

Community 0xttss Sub-Type	action	encoding
TBD3	traffic-marking-apn (Section 7.1)	4/16-octet APN ID 1-octet IPv6 (ExH) Type 1-octet Reserved
TBD4	traffic-marking-apn-partial (Section 7.2)	4/16-octet Bitmask 4/16-octet APN ID 1-octet IPv6 (ExH) Type 1-octet Reserved
TBD5	inherit-apn (Section 7.3)	4/16-octet Bitmask 1-octet IPv6 (ExH) Type 1-octet Reserved
TBD6	stitch-apn (Section 7.2)	4/16-octet Bitmask 4/16-octet APN ID 1-octet IPv6 (ExH) Type 1-octet Reserved

7.1. Traffic Marking (traffic-marking-apn) Sub-Type TBD3

The traffic-marking-apn Extended Community instructs a system to create the APN ID and encapsulate it in the indicated outer tunnel header of a transiting IP packet.

In this case, the tunnel encapsulation header is IPv6, possibly followed by an extension header (ExH). The corresponding Extended Community [RFC5701] is encoded as follows:



APN ID: 4/16 octets, APN ID value to be created and encapsulated in the indicated outer tunnel header of the transiting IP packet.

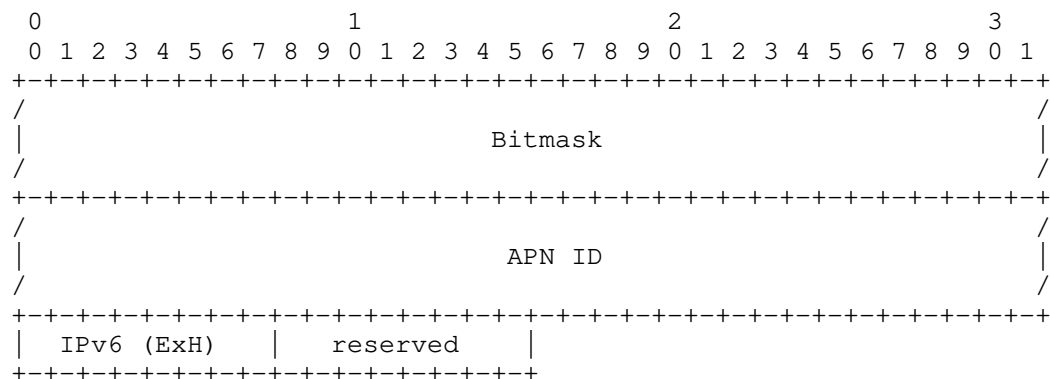
IPv6 (ExH): 1 octet, the type of each IPv6 extension header [RFC8200][RFC2780][RFC5871] is directly reused to indicate the outer tunnel to be used to encapsulate the APN ID.

reserved: 1 octet, MUST be set to 0 on encoding and MUST be ignored during decoding.

7.2. Traffic Marking (traffic-marking-apn-partial) Sub-Type TBD4

The traffic-marking-apn-partial Extended Community instructs a system to use the bitmask indicating the bits of the APN ID to be encapsulated in the indicated outer tunnel header of a transiting IP packet. The ultimately constructed APN ID may comprise of several parts obtained by the matches of different rules, and it is encapsulated in the indicated outer tunnel header.

In this case, the tunnel encapsulation header is IPv6, possibly followed by an extension header (ExH). The corresponding Extended Community [RFC5701] is encoded as follows:



Bitmask: 4/16 octets, the same length as the APN ID, indicating the bits of the APN ID to be encapsulated in the indicated outer tunnel header.

APN ID: 4/16 octets, the APN ID value to be created and encapsulated in the indicated outer tunnel header of the transiting IP packet.

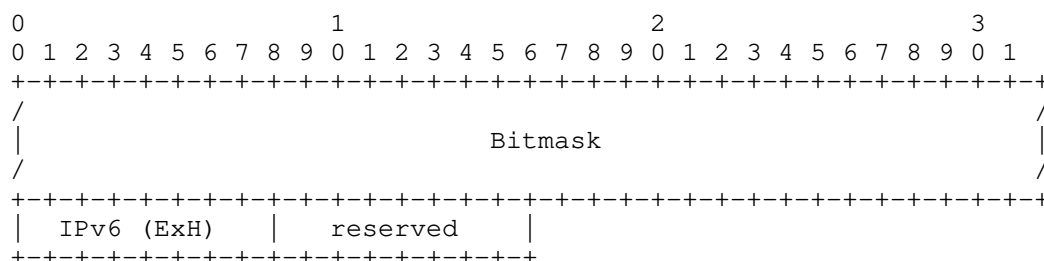
IPv6 (ExH): 1 octet, the type of each IPv6 extension header [RFC8200][RFC2780][RFC5871] is directly reused to indicate the outer tunnel to be used to encapsulate the APN ID.

reserved: 1 octet, MUST be set to 0 on encoding and MUST be ignored during decoding.

7.3. Inherit (inherit-apn) Sub-Type TBD5

The inherit-apn Extended Community instructs a system to use the Bitmask to "and" operate on the existing APN ID of a transiting IP packet and encapsulate the inherited APN ID in the indicated outer tunnel header.

In this case, the tunnel encapsulation header is IPv6, possibly followed by an extension header (ExH). The corresponding Extended Community [RFC5701] is encoded as follows:



Bitmask: 4/16 octets, the same length as the APN ID, to "and" operate on the existing APN ID of a transiting IP packet.

IPv6 (ExH): 1 octet, the type of each IPv6 extension header [RFC8200][RFC2780][RFC5871] is directly reused to indicate the outer tunnel to be used to encapsulate the APN ID.

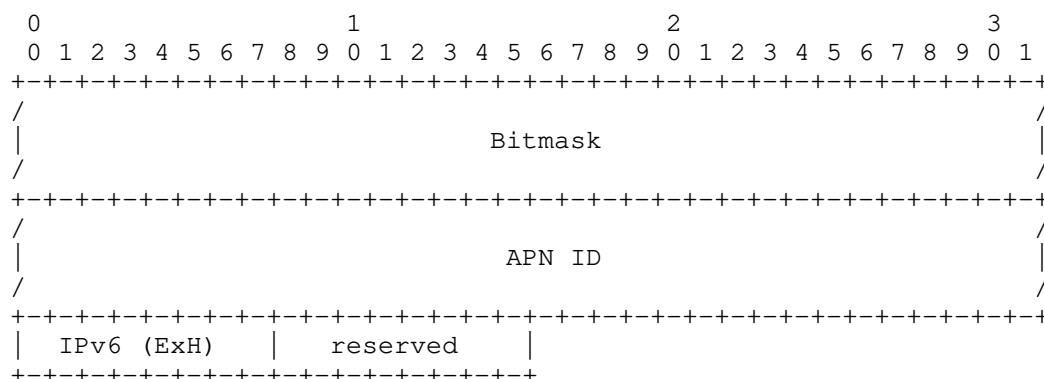
reserved: 1 octet, MUST be set to 0 on encoding and MUST be ignored during decoding.

7.4. Stitch (stitch-apn) Sub-Type TBD6

The stitch-apn Extended Community instructs a system to "and" the Bitmask with the existing APN ID of a transiting IP packet to get the part to be further encapsulated, and "and" the negation of the Bitmask with the APN ID in the Flow Spec and get the other part to be further encapsulated. The stitched APN ID is encapsulated in the indicated outer tunnel header. That is to say, the Bitmask specifies

the bits of the received APN ID to be replaced by the corresponding bits from the APN ID in the action sub-TLV value to produce a new outer APN ID. The other bits of the received APN ID are copied to the new outer AP ID.

In this case, the tunnel encapsulation header is IPv6, possibly followed by an extension header (ExH). The corresponding Extended Community [RFC5701] is encoded as follows:



Bitmask: 4/16 octets, the same length as the APN ID, used to operate on the APN ID (both carried in the transiting IP packet and in the Flow Spec).

APN ID: 4/16 octets, the APN ID value to be created and encapsulated in the indicated outer tunnel header of the transiting IP packet.

IPv6 (ExH): 1 octet, the type of each IPv6 extension header [RFC8200][RFC2780][RFC5871] is directly reused to indicate the outer tunnel to be used to encapsulate the APN ID.

reserved: 1 octet, MUST be set to 0 on encoding and MUST be ignored during decoding.

8. IANA Considerations

8.1. Flow Spec Component - APN ID

IANA is requested to assign a value in the Flow Specification Component Types Registry as follows:

Value	Name	Reference
TBD1	APN ID	This document

8.2. Opaque Extended Community - Grouping Identifier

The Grouping Identifier Opaque Extended Community is defined in this document and it is requested that a Sub-Type = TBD2 be assigned as follows.

Value	Name	Reference
TBD2	Grouping Identifier	This document

8.3. Extended Community Flow Specification Actions

The Extended Community Flow Specification Actions are defined in this document and it is requested that corresponding Sub-Types as shown in the following table be assigned.

Sub-Type Value	Name	Reference
TBD3	traffic-marking-apn	This document
TBD4	traffic-marking-apn-partial	This document
TBD5	inherit-apn	This document
TBD6	stitch-apn	This document

9. Acknowledgements

The authors would like to thank the careful reviews and valuable comments from Haibo Wang, Shunwan Zhuang, Stefano Previdi, and Donald Eastlake.

10. Security Considerations

The security considerations are the same as [RFC8955], [RFC8956], and [I-D.li-apn-framework].

11. References

11.1. Normative References

[I-D.hares-idr-flowspec-v2]

Hares, S., Eastlake, D., Yadlapalli, C., and S. Maduschke, "BGP Flow Specification Version 2", Work in Progress, Internet-Draft, draft-hares-idr-flowspec-v2-05, 4 February 2022, <<https://www.ietf.org/internet-drafts/draft-hares-idr-flowspec-v2-05.txt>>.

[I-D.li-apn-framework]

Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C., and G. Mishra, "Application-aware Networking (APN) Framework", Work in Progress, Internet-Draft, draft-li-apn-framework-05, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-framework-05.txt>>.

[I-D.li-apn-header]

Li, Z., Peng, S., and S. Zhang, "Application-aware Networking (APN) Header", Work in Progress, Internet-Draft, draft-li-apn-header-02, 7 April 2022, <<https://www.ietf.org/archive/id/draft-li-apn-header-02.txt>>.

[I-D.li-apn-ipv6-encap]

Li, Z., Peng, S., and C. Xie, "Application-aware IPv6 Networking (APN6) Encapsulation", Work in Progress, Internet-Draft, draft-li-apn-ipv6-encap-04, 7 April 2022, <<https://www.ietf.org/archive/id/draft-li-apn-ipv6-encap-04.txt>>.

[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>>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <<https://www.rfc-editor.org/info/rfc4360>>.

- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <<https://www.rfc-editor.org/info/rfc4760>>.
- [RFC5701] Rekhter, Y., "IPv6 Address Specific BGP Extended Community Attribute", RFC 5701, DOI 10.17487/RFC5701, November 2009, <<https://www.rfc-editor.org/info/rfc5701>>.
- [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>>.
- [RFC8955] Loibl, C., Hares, S., Raszuk, R., McPherson, D., and M. Bacher, "Dissemination of Flow Specification Rules", RFC 8955, DOI 10.17487/RFC8955, December 2020, <<https://www.rfc-editor.org/info/rfc8955>>.
- [RFC8956] Loibl, C., Ed., Raszuk, R., Ed., and S. Hares, Ed., "Dissemination of Flow Specification Rules for IPv6", RFC 8956, DOI 10.17487/RFC8956, December 2020, <<https://www.rfc-editor.org/info/rfc8956>>.

11.2. Informative References

- [I-D.ietf-idr-flowspec-l2vpn] Hao, W., Eastlake, D. E., Litkowski, S., and S. Zhuang, "BGP Dissemination of L2 Flow Specification Rules", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-l2vpn-19, 18 April 2022, <<https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-l2vpn-19.txt>>.
- [I-D.li-apn-problem-statement-usecases] Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z., and G. Mishra, "Problem Statement and Use Cases of Application-aware Networking (APN)", Work in Progress, Internet-Draft, draft-li-apn-problem-statement-usecases-06, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-problem-statement-usecases-06.txt>>.
- [RFC2780] Bradner, S. and V. Paxson, "IANA Allocation Guidelines For Values In the Internet Protocol and Related Headers", BCP 37, RFC 2780, DOI 10.17487/RFC2780, March 2000, <<https://www.rfc-editor.org/info/rfc2780>>.
- [RFC5871] Arkko, J. and S. Bradner, "IANA Allocation Guidelines for the IPv6 Routing Header", RFC 5871, DOI 10.17487/RFC5871, May 2010, <<https://www.rfc-editor.org/info/rfc5871>>.

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

Authors' Addresses

Shuping Peng
Huawei Technologies
Beijing
China
Email: pengshuping@huawei.com

Zhenbin Li
Huawei Technologies
Beijing
China
Email: lizhenbin@huawei.com

Sheng Fang
Huawei Technologies
Beijing
China
Email: fangsheng@huawei.com

Yong Cui
Tsinghua University
Beijing
China
Email: cuiyong@tsinghua.edu.cn

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 8 September 2022

S. Peng
Z. Li
Huawei Technologies
G. Mishra
Verizon Inc.
7 March 2022

APN Scope and Gap Analysis
draft-peng-apn-scope-gap-analysis-04

Abstract

The APN work in IETF is focused on developing a framework and set of mechanisms to derive, convey and use an attribute allowing the implementation of fine-grain user group-level and application group-level requirements in the network layer. APN aims to apply various policies in different nodes along a network path onto a traffic flow altogether, for example, at the headend to steer into corresponding path, at the midpoint to collect corresponding performance measurement data, and at the service function to execute particular policies. Currently there is still no way to efficiently realize this composite network service provisioning along the path. This document further clarifies the scope of the APN work and describes the solution gap analysis.

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 8 September 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	2
2. Requirements Language	3
3. Terminologies	3
4. APN Framework and Scope	3
5. Example Use Case and Existing Issues	4
6. Basic Solution and Benefits	5
7. Solution Gap Analysis	7
7.1. IPv6/MPLS Flow Label	7
7.2. SFC ServiceID	7
7.3. IOAM Flow ID	8
7.4. Binding SID	9
7.5. FlowSpec Label	9
7.6. Group Policy ID	9
7.7. Detnet Flow Identification	9
7.8. Network Slicing Resource ID	10
7.9. Service Path ID	10
7.10. Summary	10
8. IANA Considerations	11
9. Acknowledgements	11
10. Informative References	11
Authors' Addresses	15

1. Introduction

Application-aware Networking (APN) is introduced in [I-D.li-apn-framework] and [I-D.li-apn-problem-statement-usecases]. APN conveys an attribute along with data packets into network and makes the network aware about data flow requirements at different granularity levels.

Such an attribute is acquired, constructed in a structured value, and then encapsulated in the packet. Such structured value is treated as an opaque object in the network to which the network operator applies policies in various nodes/service functions along the path and provides corresponding services.

This structured attribute can be encapsulated in various data planes adopted within a Network Operator controlled limited domain, e.g. MPLS, VXLAN, SR/SRv6 and other tunnel technologies, which waits to be further specified.

With APN, it becomes possible to apply various policies in different nodes along a network path onto a traffic flow altogether in a more efficient way, e.g., at the headend to steer into corresponding path, at the midpoint to collect corresponding performance measurement data, and at the service function to execute particular policies. Currently there is still no way to realize this composite network service provisioning along the path very efficiently. It may be possible to stack those various policies in a list of TLVs at the headend. However, this approach would introduce great complexities and impose big challenges on the hardware processing and forwarding.

The example use-case presented in this draft further expands on the rationale for such an attribute and how it can be derived and used in that specific context.

This document further clarifies the scope of the APN work and describes the solution gap analysis.

2. Requirements Language

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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminologies

APN: Application-aware Networking

CPE: Customer Premises Equipment

DPI: Deep Packet Inspection

OS: Operating System

4. APN Framework and Scope

The APN framework is introduced in [I-D.li-apn-framework], as shown in the Figure 1.

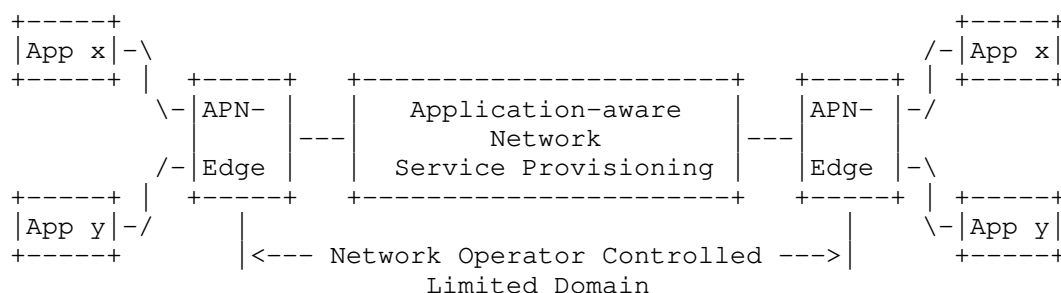


Figure 1. APN Framework and Scope

APN is only applied to an edge-to-edge tunnel encapsulation within a limited trusted domain. It means that the source and destination addresses of the packet are the endpoints of the tunnel (i.e. the domain edges), and nothing about the payload source and destination can be deduced, which substantially reduces the privacy concerns. Typically, an APN domain is defined as a Network Operator controlled limited domain (see Figure 1), in which MPLS, VXLAN, SR/SRv6 and other tunnel technologies are adopted to provide network services.

With APN, the attribute is acquired based on the existing information in the packet header (i.e. source and destination addresses, incoming L2 (or) MPLS encapsulation, incoming physical/virtual port information, the other fields of the 5-tuple if they are not encrypted) at the edge devices of the APN domain, added to the data packets along with the tunnel encapsulation, and delivered to the network, wherein, according to this attribute, corresponding network services are provisioned. When the packets leave the APN domain, the attribute is removed together with the tunnel encapsulation header.

5. Example Use Case and Existing Issues

To be more specific and more concrete, here we use SD-WAN as an example use case to further expand on the rationale for such attribute and how it can be derived and used in that specific context.

In the case of SD-WAN, an enterprise obtains WAN services from an SD-WAN provider so that its employees have access to the applications in the Cloud, and then the SD-WAN provider may buy WAN lines from a Network Operator. The enterprise may know what applications will use the SD-WAN services, but it will only provide the 5 tuples (i.e. source IP address, source port, destination IP address, destination port, transport protocol) of those applications to the SD-WAN

provider. So, the SD-WAN provider does not know what applications it is serving, and will only provide 5 tuples to the Network Operator and the service performance requirements for steering their customer's traffic. In this way, the Network Operator does not know anything else about the traffic except the 5 tuples and requirements. Nowadays, SD-WAN is usually using 5-tuple to steer the traffic into corresponding WAN lines across the Network Operator's network [SD-WAN].

However, there are two main issues in the current SD-WAN deployments.

1) It is complicated to resolve the 5 tuples. Even worse, as the traffic is encrypted, it becomes impossible to obtain any transport layer information. Moreover, in the IPv6 data plane, with the extension headers being added before the upper layer, in some implementations it becomes very difficult and even impossible to obtain transport layer information because that information is located deep in the packet. So, there is no 5 tuples anymore, and maybe only 2 tuples are available.

2) Currently there is still no way to apply various policies in different nodes along the network path onto a traffic flow altogether, that is, at the headend to steer into corresponding path, at the midpoint to collect corresponding performance measurement data, and at the service function to execute particular policies. It may be possible to stack those various policies in a list of TLVs at the headend. However, this approach would introduce great complexities and impose big challenges on the hardware processing and forwarding.

6. Basic Solution and Benefits

With APN, at the edge node, i.e. CPE, of the SD-WAN (see Figure 2), the 5-tuple, plus information related to user or application group-level requirements is constructed into a structured value, called APN attribute. This attribute is only meaningful for the network operators to apply various policies in different nodes/service functions, which can be enforced from the Controllers.

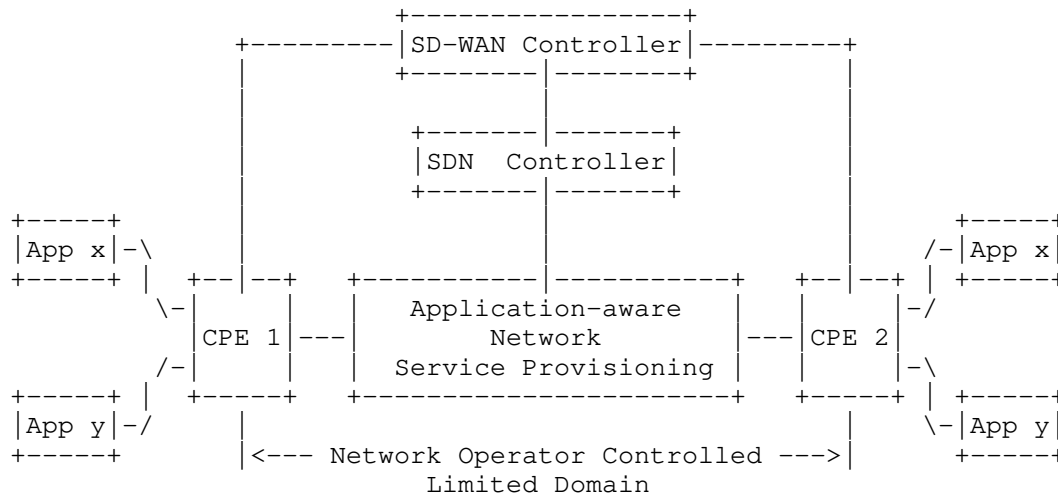


Figure 2. SD-WAN using the APN Framework

With such an attribute in the network, we can easily solve the two issues above-mentioned. For example, when the packet is sent from the CPE1 and the attribute is added along with the tunnel encapsulation, then it is not necessary to resolve the 5-tuple and perform the deep inspection in every node along the path. This attribute is encapsulated in the network layer and can be easily read by the routers and service functions. If the tunnel is based on the IPv6 data plane, for example, such an attribute can be encapsulated in an option of IPv6 hop-by-hop options header.

Since this attribute is taken as an object to the network, the network operators will simply place the policies in the nodes/service functions where this indicated traffic will go through, and the corresponding node/service function will just apply policies for this object. This can be easily done by utilizing this attribute, which is not possible with any current existing mechanism.

Such attribute will also bring other benefits, for example,

- * Improve the forwarding performance since it will only use 1 field in the IP layer instead of resolving 5 tuples, which will also improve the scalability.
- * Very flexible policy enforcement in various nodes and service functions along the network path.

Furthermore, with such attribute, more new services could be enabled, for example,

- * Even more fine-granularity performance measurement could be achieved and the granularity to be monitored and visualized can be controllable, which is able to relieve the processing pressure on the controller when it is facing the massive monitoring data.
- * The policy execution on the service function can be based only on this value and not based on 5-tuple, which can eliminate the need of deep packet inspection.
- * The underlay performance guarantee could be achieved for SD-WAN overlay services, such as explicit traffic engineering path satisfying SLA and selective visualized accurate performance measurement.

7. Solution Gap Analysis

There are already some solutions specified in IETF, which use identifier to perform traffic steering and service provisioning. However, the existing solutions are specific to a particular scenario or data plane. None of them is the same as APN and able to achieve the same effects.

7.1. IPv6/MPLS Flow Label

[RFC6437] specifies the IPv6 flow label which enables the IPv6 flow classification. However, the IPv6 flow label is mainly used for Equal Cost Multipath Routing (ECMP) and Link Aggregation [RFC6438].

Similarly, [RFC6391] describes a method of adding an additional Label Stack Entry (LSE) at the bottom of the stack in order to facilitate the load balancing of the flows within a pseudowire (PW) over the available ECMPs. A similar design for general MPLS use has also been proposed in [RFC6790] using the concept of Entropy Label.

7.2. SFC ServiceID

Subscriber Identifier and Performance Policy Identifier are specified in [RFC8979]. These identifiers are carried only in the Network Service Header (NSH) [RFC8300] Context Header, as shown in Figure 3, while the APN attribute can be carried in various data plane encapsulations.

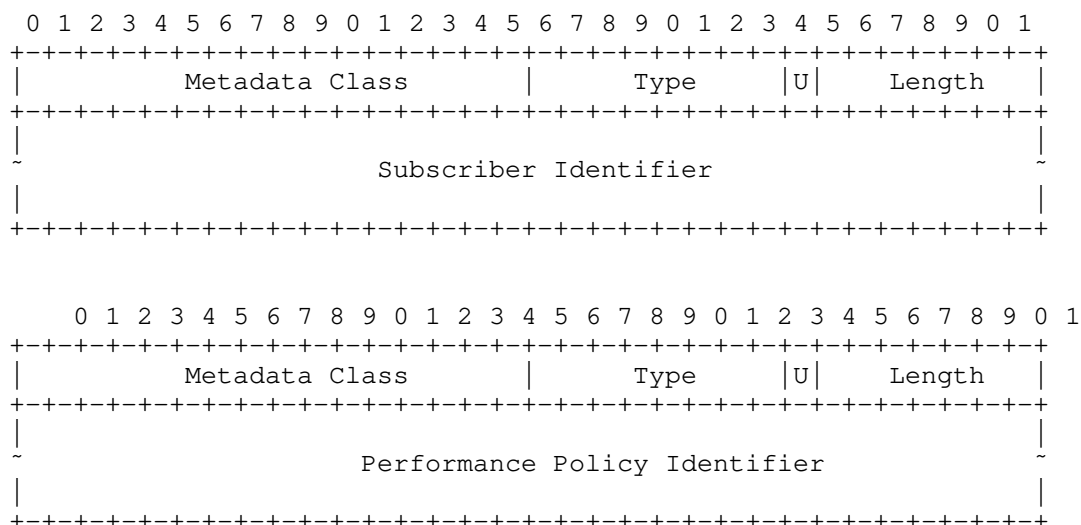


Figure 3. Subscriber Identifier and Performance Policy Identifier

In this draft [RFC8979], the Subscriber Identifier carries an opaque local identifier that is assigned to a subscriber by a network operator, and the Performance Policy Identifier represents an opaque value pointing to specific performance policy to be enforced. In this way, in order to apply various policies in different nodes along the network path onto a traffic flow altogether, e.g., at the headend to steer into corresponding path, at the midpoint to collect corresponding performance measurement data, and at the service function to execute particular policies, those various policies would have to be stacked in a list of TLVs at the headend, introducing great complexities and big challenges on the hardware processing and forwarding.

The APN attribute is treated as an opaque object in the network, to which the network operator applies policies in various nodes/service functions along the path and provide corresponding services.

7.3. IOAM Flow ID

A 32-bit Flow ID is specified in [I-D.ietf-ippm-ioam-direct-export], which is used to correlate the exported data of the same flow from multiple nodes and from multiple packets, while the APN attribute can serve more various purposes.

7.4. Binding SID

The Binding SID (BSID) [RFC8402] is bound to an SR Policy, instantiation of which may involve a list of SIDs. Any packets received with an active segment equal to BSID are steered onto the bound SR Policy. A BSID may be either a local or a global SID. While the APN attribute is not bound to SR only, and it can be carried in various data plane encapsulations.

7.5. FlowSpec Label

The flow specification (FlowSpec) [RFC5575] is actually an n-tuple consisting of several matching criteria that can be applied to IP traffic, which include elements such as source and destination address prefixes, IP protocol, and transport protocol port numbers. In BGP VPN/MPLS networks, BGP FlowSpec can be extended to identify and change (push/swap/pop) the label(s) for traffic that matches a particular FlowSpec rule in [I-D.ietf-idr-flowspec-mpls-match] and [I-D.ietf-idr-bgp-flowspec-label]. In [I-D.liang-idr-bgp-flowspec-route], BGP is used to distribute the FlowSpec rule bound with label(s). While the APN attribute is not bound to MPLS only, and it can be carried in various data plane encapsulations.

7.6. Group Policy ID

The capabilities of the VXLAN-GPE protocol can be extended by defining next protocol "shim" headers that are used to implement new data plane functions. For example, Group Policy ID is carried in the Group-Based Policy (GBP) Shim header [I-D.lemon-vxlan-lisp-gpe-gbp]. GENEVE has similar ability as VXLAN-GPE to carry metadata.

7.7. Detnet Flow Identification

Identification and Specification of DetNet Flows is specified in [RFC9016]. DetNet MPLS flows can be identified and specified by the SLabel and the FLabelStack. The IP 6-tuple is used for DetNet IP flow identification, which consists of SourceIpAddress, DestinationIpAddress, Dscp, Protocol, SourcePort, and DestinationPort. IPv6FlowLabel and IPsecSpi are additional attributes that can be used for DetNet flow identification in addition to the 6-tuple. Therefore, the Detnet IP Flow ID is logical and there is no such Flow ID carried for Detnet, but only the 6-tuple is directly used to identify the Detnet flows.

Only one exceptional case, in [I-D.ietf-spring-sr-redundancy-protection], the 32-bit flow identification (FID) identifies one specific Detnet flow of

redundancy protection. This FID is usually allocated from centralized controller to the SR ingress node or redundancy node in SR network.

7.8. Network Slicing Resource ID

In [I-D.dong-6man-enhanced-vpn-vtn-id], VTN Resource ID is a 4-octet identifier which uniquely identifies the set of network resources allocated to a VTN. For network slicing, the ID is used to indicate the network resources to be allocated to the network slices and it is not bound to any traffic flow.

APN is for traffic steering, while network slicing is about resource partition [I-D.ietf-teas-rfc3272bis].

7.9. Service Path ID

In [RFC8300], Service Path Identifier (SPI) uniquely identifies a Service Function Path (SFP). Participating nodes MUST use this identifier for SFP selection. The initial Classifier MUST set the appropriate SPI for a given classification result. For SFC, the ID is used to indicate a SF path and it is not bound to any traffic flow.

7.10. Summary

The comparison of the identifiers for the typical network services (incl. iOAM, Detnet, Network Slicing (NS), and Service Function Chaining (SFC)) is shown in the following Table from different aspects (incl. ID, Identification Object, Source (for generating the ID), Configuration (Conf.) node, and Size).

	ID	Identification Object	Source	Conf. node	Size
APN	APN ID	The flow that needs fine-granular services	5-tuple Layer 2	Controller	32bits 128b
iOAM	Flow ID	The flow that needs performance monitoring	-	Controller Ingress	32bits
Detnet	Flow ID (6-tuple)	The flow that needs Detnet services	-	Controller	-
Detnet	Flow ID	The redundant protection flow	-	Detnet Controller	32bits
NS	Resource ID	The network resources that are allocated to network slices	-	Controller	32bits
SFC	SPI	The SF Path	-	Controller	24bits
SFC	Performance Policy ID	The performance policy	-	Controller	-

Table 1. Comparison of the Identifiers

As driven by ever-emerging new 5G services, fine-granularity service provisioning becomes urgent. The existing solutions are either specific to a particular scenario or data plane. While APN aims to define a generalized attribute used for fine-granularity service provisioning, and can be carried in various data plane encapsulations.

8. IANA Considerations

There are no IANA considerations in this document.

9. Acknowledgements

The authors would like to acknowledge Martin Vigoureux, Alvaro Retana, Barry Leiba, Stefano Previdi, Adrian Farrel, and Daniel King for their valuable review and comments.

10. Informative References

[I-D.brockners-ippm-ioam-vxlan-gpe]

Brockners, F., Bhandari, S., Govindan, V. P., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mizrahi, T., Kfir, A., Gafni, B., Lapukhov, P., and M. Spiegel, "VXLAN-GPE Encapsulation for In-situ OAM Data", Work in Progress, Internet-Draft, draft-brockners-ippm-ioam-vxlan-gpe-03, 4 November 2019, <<https://www.ietf.org/archive/id/draft-brockners-ippm-ioam-vxlan-gpe-03.txt>>.

[I-D.dong-6man-enhanced-vpn-vtn-id]

Dong, J., Li, Z., Xie, C., Ma, C., and G. Mishra, "Carrying Virtual Transport Network (VTN) Identifier in IPv6 Extension Header", Work in Progress, Internet-Draft, draft-dong-6man-enhanced-vpn-vtn-id-06, 24 October 2021, <<https://www.ietf.org/archive/id/draft-dong-6man-enhanced-vpn-vtn-id-06.txt>>.

[I-D.ietf-idr-bgp-flowspec-label]

Liang, Q., Hares, S., You, J., Raszuk, R., and D. Ma, "Carrying Label Information for BGP FlowSpec", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-flowspec-label-01, 6 December 2016, <<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-flowspec-label-01.txt>>.

[I-D.ietf-idr-flowspec-mpls-match]

Yong, L., Hares, S., Liang, Q., and J. You, "BGP Flow Specification Filter for MPLS Label", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-mpls-match-01, 6 December 2016, <<https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-mpls-match-01.txt>>.

[I-D.ietf-ippm-ioam-direct-export]

Song, H., Gafni, B., Zhou, T., Li, Z., Brockners, F., Bhandari, S., Sivakolundu, R., and T. Mizrahi, "In-situ OAM Direct Exporting", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-direct-export-07, 13 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-direct-export-07.txt>>.

[I-D.ietf-sfc-serviceid-header]

Sarikaya, B., Hugo, D. V., and M. Boucadair, "Subscriber and Performance Policy Identifier Context Headers in the Network Service Header (NSH)", Work in Progress, Internet-Draft, draft-ietf-sfc-serviceid-header-14, 11 December 2020, <<https://www.ietf.org/archive/id/draft-ietf-sfc-serviceid-header-14.txt>>.

- [I-D.ietf-spring-sr-redundancy-protection]
Geng, X., Chen, M., Yang, F., Garvia, P. C., and G. Mishra, "SRv6 for Redundancy Protection", Work in Progress, Internet-Draft, draft-ietf-spring-sr-redundancy-protection-01, 15 February 2022, <<https://www.ietf.org/archive/id/draft-ietf-spring-sr-redundancy-protection-01.txt>>.
- [I-D.ietf-teas-rfc3272bis]
Farrel, A., "Overview and Principles of Internet Traffic Engineering", Work in Progress, Internet-Draft, draft-ietf-teas-rfc3272bis-15, 24 February 2022, <<https://www.ietf.org/archive/id/draft-ietf-teas-rfc3272bis-15.txt>>.
- [I-D.lemon-vxlan-lisp-gpe-gbp]
Lemon, J., Maino, F., Smith, M., and A. Isaac, "Group Policy Encoding with VXLAN-GPE and LISP-GPE", Work in Progress, Internet-Draft, draft-lemon-vxlan-lisp-gpe-gbp-02, 30 April 2019, <<https://www.ietf.org/archive/id/draft-lemon-vxlan-lisp-gpe-gbp-02.txt>>.
- [I-D.li-6man-app-aware-ipv6-network]
Li, Z., Peng, S., Li, C., Xie, C., Voyer, D., Li, X., Liu, P., Cao, C., and K. Ebisawa, "Application-aware IPv6 Networking (APN6) Encapsulation", Work in Progress, Internet-Draft, draft-li-6man-app-aware-ipv6-network-03, 22 February 2021, <<https://www.ietf.org/archive/id/draft-li-6man-app-aware-ipv6-network-03.txt>>.
- [I-D.li-apn-framework]
Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C., Mishra, G., Ebisawa, K., Previdi, S., and J. N. Guichard, "Application-aware Networking (APN) Framework", Work in Progress, Internet-Draft, draft-li-apn-framework-04, 25 October 2021, <<https://www.ietf.org/archive/id/draft-li-apn-framework-04.txt>>.
- [I-D.li-apn-problem-statement-usecases]
Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z., Mishra, G., Ebisawa, K., Previdi, S., and J. N. Guichard, "Problem Statement and Use Cases of Application-aware Networking (APN)", Work in Progress, Internet-Draft, draft-li-apn-problem-statement-usecases-05, 20 December 2021, <<https://www.ietf.org/archive/id/draft-li-apn-problem-statement-usecases-05.txt>>.

- [I-D.liang-idr-bgp-flowspec-route]
Liang, Q. and J. You, "BGP FlowSpec based Multi-dimensional Route Distribution", Work in Progress, Internet-Draft, draft-liang-idr-bgp-flowspec-route-00, 20 October 2014, <<https://www.ietf.org/archive/id/draft-liang-idr-bgp-flowspec-route-00.txt>>.
- [I-D.peng-apn-security-privacy-consideration]
Peng, S., Li, Z., Voyer, D., Li, C., Liu, P., and C. Cao, "APN Security and Privacy Considerations", Work in Progress, Internet-Draft, draft-peng-apn-security-privacy-consideration-02, 16 June 2021, <<https://www.ietf.org/archive/id/draft-peng-apn-security-privacy-consideration-02.txt>>.
- [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>>.
- [RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <<https://www.rfc-editor.org/info/rfc5575>>.
- [RFC6391] Bryant, S., Ed., Filsfils, C., Drafz, U., Kompella, V., Regan, J., and S. Amante, "Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network", RFC 6391, DOI 10.17487/RFC6391, November 2011, <<https://www.rfc-editor.org/info/rfc6391>>.
- [RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, DOI 10.17487/RFC6437, November 2011, <<https://www.rfc-editor.org/info/rfc6437>>.
- [RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011, <<https://www.rfc-editor.org/info/rfc6438>>.
- [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>>.

- [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>>.
- [RFC8300] Quinn, P., Ed., Elzur, U., Ed., and C. Pignataro, Ed., "Network Service Header (NSH)", RFC 8300, DOI 10.17487/RFC8300, January 2018, <<https://www.rfc-editor.org/info/rfc8300>>.
- [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>>.
- [RFC8979] Sarikaya, B., von Hugo, D., and M. Boucadair, "Subscriber and Performance Policy Identifier Context Headers in the Network Service Header (NSH)", RFC 8979, DOI 10.17487/RFC8979, February 2021, <<https://www.rfc-editor.org/info/rfc8979>>.
- [RFC9016] Varga, B., Farkas, J., Cummings, R., Jiang, Y., and D. Fedyk, "Flow and Service Information Model for Deterministic Networking (DetNet)", RFC 9016, DOI 10.17487/RFC9016, March 2021, <<https://www.rfc-editor.org/info/rfc9016>>.
- [SD-WAN] MEF 70.1 Draft (R1), available at <https://www.mef.net/wp-content/uploads/2020/08/MEF-70-1-Draft-R1.pdf>/, "SD-WAN Service Attributes and Service Framework", August 2020.

Authors' Addresses

Shuping Peng
Huawei Technologies
Beijing
China
Email: pengshuping@huawei.com

Zhenbin Li
Huawei Technologies
Beijing
China
Email: lizhenbin@huawei.com

Gyan Mishra
Verizon Inc.
United States of America
Email: gyan.s.mishra@verizon.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 31 October 2022

S. Peng
Z. Li
Huawei Technologies
29 April 2022

A YANG Model for Application-aware Networking (APN)
draft-peng-apn-yang-01

Abstract

Application-aware Networking (APN) is a framework, where APN data packets convey APN attribute (incl. APN ID and/or APN Parameters) to enable fine grained service provisioning. This document defines a YANG module for APN.

The YANG modules in this document conform to the Network Management Datastore Architecture (NMDA).

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 31 October 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	2
2. Terminologies	3
3. APN Configuration data model	3
3.1. APN YANG Model Structure	3
3.2. APN ID Template	5
3.3. APN ID Marking	5
3.4. APN Policy Mapping	6
4. APN YANG Module	7
5. IANA Considerations	16
6. Security Considerations	17
7. Acknowledgements	17
8. Normative References	17
Authors' Addresses	18

1. Introduction

Application-aware Networking (APN) is introduced in [I-D.li-apn-framework] and [I-D.li-apn-problem-statement-usecases]. APN data packets convey the APN attribute (incl. APN ID and/or APN Parameters). The APN ID is a structured value, treated as an opaque object in the network, to which the network operator applies policies in various nodes/service functions along the path so to provide corresponding services. In an IPv6 network, a design proposal of such structured value can refer to [I-D.li-apn-header]. The APN attribute can be encapsulated in various data plane adopted within a Network Operator controlled limited domain, e.g. IPv6, MPLS, and other tunnel technologies, which wait to be further specified.

This document defines a data model for APN using the YANG data modeling language [RFC7950]. This YANG model supports the APN Attribute options [I-D.li-apn-framework].

The modeling in this document complies with the Network Management Datastore Architecture (NMDA) defined in [RFC8342].

2. Terminologies

APN: Application-aware Networking

APN ID: APN Identifier

The terminology for describing YANG data models is found in [RFC7950].

Tree diagrams used in this document follow the notation defined in [RFC8340].

3. APN Configuration data model

3.1. APN YANG Model Structure

The APN YANG model includes the data plane protocol indication, the global actions, the apn-id-template configuration, the apn-id-marking, and the mapping policies for APN. The structure of the APN YANG model is shown in Figure 1.

The APN YANG model can cover several data plane protocols. In this model, only IPv6 is presented.

One global action is defined currently, i.e., the apn-id-inherit, which is used to configure the APN ID carried in the inner packet to be inherited (copied and encapsulated) into the outer tunnel header.

The apn-id-templates configures the templates of the APN ID. More than one templates can be configured.

The apn-id-marking configures the APN ID on the flow which is identified by the selected filter.

The mapping-policies configures the APN ID based on the selected template, and the to-be-mapped-into policy based on the configured APN ID. More than one policies can be configured.

```
module: ietf-apn
  +--rw apn!
    +--rw ipv6!
      +--rw global
        | +--rw apn-id-inherit?    apn-id-inherit-type
      +--rw apn-id-templates
        | +--rw apn-id-template* [name]
        |   +--rw name              string
        |   +--rw app-info-fields!
```

```

    +---rw app-fields
    |   +---rw app-field* [index]
    |   |   +---rw index      uint32
    |   |   +---rw name       string
    |   |   +---rw length?    uint32
    +---rw user-info-fields!
    |   +---rw user-fields
    |   |   +---rw user-field* [index]
    |   |   |   +---rw index      uint32
    |   |   |   +---rw name       string
    |   |   |   +---rw length?    uint32
+---rw apn-id-marking!
|   +---rw filter
|   |   +---rw filter-type?    apn-filter-type
|   |   +---rw ace-name?       -> /acl:acls/acl/aces/ace/name
+---rw apn-ipv6-template      -> /apn/ipv6/apn-id-templates/apn-id-templa
te/name
|   +---rw app-fields
|   |   +---rw app-field* [name]
|   |   |   +---rw name       -> /apn/ipv6/apn-id-templates/apn-id-template[apn
:name=current()/../../../../apn-ipv6-template]/app-info-fields/app-fields/app-field/
name
|   |   |   +---rw value      uint32
+---rw user-fields
|   +---rw user-field* [name]
|   |   +---rw name       -> /apn/ipv6/apn-id-templates/apn-id-template[apn
:name=current()/../../../../apn-ipv6-template]/user-info-fields/user-fields/user-fie
ld/name
|   |   +---rw value      uint32
+---rw mapping-policys
|   +---rw mapping-policy* [color]
|   |   +---rw color      uint32
|   |   +---rw name       string
|   |   +---rw description? string
+---rw apn-id-template?      -> /apn/ipv6/apn-id-templates/apn-id-templ
ate/name
|   +---rw apn-ipv6-maps
|   |   +---rw apn-ipv6-map* [index]
|   |   |   +---rw index      uint32
|   |   |   +---rw app-fields
|   |   |   |   +---rw app-field* [name]
|   |   |   |   |   +---rw name       -> /apn/ipv6/apn-id-templates/apn-id-tem
plate[apn:name=current()/../../../../../../../../apn-id-template]/app-info-fields/app-fiel
ds/app-field/name
|   |   |   |   |   +---rw value      uint32
+---rw user-fields
|   |   |   |   +---rw user-field* [name]
|   |   |   |   |   +---rw name       -> /apn/ipv6/apn-id-templates/apn-id-tem
plate[apn:name=current()/../../../../../../../../apn-id-template]/user-info-fields/user-fi
elds/user-field/name
|   |   |   |   |   +---rw value      uint32
+---rw (match-tunnel)
|   |   |   |   +---:(sr-policy)
|   |   |   |   |   +---rw color?      uint32
+---:(ip)
|   |   |   |   |   +---rw native-ip?  empty

```


Figure 1. APN YANG Model Structure

3.2. APN ID Template

The APN ID template can be configured with the defined fields, including the app-info-fields and the user-info-fields, each of which can have several fields with their name and length configurable.

```
+--rw apn-id-templates
  +--rw apn-id-template* [name]
    +--rw name                string
    +--rw app-info-fields!
      +--rw app-fields
        +--rw app-field* [index]
          +--rw index          uint32
          +--rw name            string
          +--rw length?         uint32
      +--rw user-info-fields!
        +--rw user-fields
          +--rw user-field* [index]
            +--rw index          uint32
            +--rw name            string
            +--rw length?         uint32
```

3.3. APN ID Marking

The APN ID Marking uses the selected filter to identify the flow on which APN is applied. Multiple filter types exist. ACL [RFC8519] is a common way to specify a flow.

Upon the identified flow, the APN template is used to configure the APN ID with the defined fields, including the app-info-fields and the user-info-fields, each of which can have several fields with their name and length configurable.

```

+---rw apn-id-marking!
  +---rw filter
    |   +---rw filter-type?    apn-filter-type
    |   +---rw ace-name?      -> /acl:acls/acl/aces/ace/name
  +---rw apn-ipv6-template    -> /apn/ipv6/apn-id-templates/apn-i
  +---rw app-fields
    |   +---rw app-field* [name]
    |   +---rw name          -> /apn/ipv6/apn-id-templates/apn-id-tempn-ipv6-t
emplate]/app-info-fields/app-fields/app-field/name
    |   +---rw value        uint32
  +---rw user-fields
    |   +---rw user-field* [name]
    |   +---rw name          -> /apn/ipv6/apn-id-templates/apn-id-tempn-ipv6-t
emplate]/user-info-fields/user-fields/user-field/name
    +---rw value            uint32

```

3.4. APN Policy Mapping

The APN policy mapping is for mapping to corresponding policies based on the APN ID being structured with the configured fields. The mapping into SR policy is presented in the model below.

```

+---rw mapping-policys
  +---rw mapping-policy* [color]
    +---rw color            uint32
    +---rw name             string
    +---rw description?     string
    +---rw apn-id-template? -> /apn/ipv6/apn-id-templates/apn-id-templ
ate/name
  +---rw apn-ipv6-maps
    +---rw apn-ipv6-map* [index]
      +---rw index          uint32
      +---rw app-fields
        |   +---rw app-field* [name]
        |   +---rw name      -> /apn/ipv6/apn-id-templates/apn-id-tem
plate[apn:name=current()/../../../../../../../../apn-id-template]/app-info-fields/app-fiel
ds/app-field/name
        |   +---rw value    uint32
      +---rw user-fields
        |   +---rw user-field* [name]
        |   +---rw name      -> /apn/ipv6/apn-id-templates/apn-id-tem
plate[apn:name=current()/../../../../../../../../apn-id-template]/user-info-fields/user-fi
elds/user-field/name
        |   +---rw value    uint32
      +---rw (match-tunnel)
        +---:(sr-policy)
          |   +---rw color?    uint32
          +---:(ip)
            +---rw native-ip?  empty

```


4. APN YANG Module

```
module ietf-apn {
  namespace "urn:ietf:params:xml:ns:yang:ietf-apn";
  prefix apn;

  import ietf-access-control-list {
    prefix "acl";
    reference
      "RFC 8519: YANG Data Model for Network Access Control
       Lists (ACLs)";
  }

  organization
    "APN";

  contact
    "Web: <https://datatracker.ietf.org/wg/apn/about/>
     WG List: <apn@ietf.org>
     Editor: pengshuping@huawei.com;

  description
    "This YANG module specifies a vendor-independent data
     model for the Application-aware Networking (APN).

    Copyright (c) 2020 IETF Trust and the persons identified as
    authors of the code. All rights reserved.

    Redistribution and use in source and binary forms, with or
    without modification, is permitted pursuant to, and subject
    to the license terms contained in, the Simplified BSD License
    set forth in Section 4.c of the IETF Trust's Legal Provisions
    Relating to IETF Documents
    (http://trustee.ietf.org/license-info).

    This version of this YANG module is part of RFC XXXX; see the
    RFC itself for full legal notices.";

  revision 2021-10-20 {
    description "Initial revision.";
    reference "draft-peng-apn-yang";
  }

  /*
   * IDENTITIES
   */
```

```
identity base-filter {
  description
    "Base identity to represent a filter. A filter is used to
    specify the flow to mark the APN ID. ";
}

identity acl-filter {
  base base-filter;
  description
    "Apply ACL rules to specify the flow.";
}

/*
 * TYPE DEFINITIONS
 */

typedef apn-id-inherit-type {
  type enumeration {
    enum "enable" {
      value 1;
      description
        "Inherit the APN ID.";
    }
    enum "disable" {
      value 2;
      description
        "Not inherit the APN ID.";
    }
  }
  description
    "APN ID inherit type.";
}

typedef template-state-type {
  type enumeration {
    enum "unavailable" {
      value 0;
      description
        "The APN ID template is unavailable.";
    }
    enum "available" {
      value 1;
      description
        "The APN ID template is available.";
    }
  }
  description
    "APN ID template state type.";
}
```

```
    }

    typedef apn-filter-type {
      type identityref {
        base base-filter;
      }
      description
        "Specifies a known type of filter.";
    }

/*
 * GROUP DEFINITIONS
 */

grouping apn-filter {
  description "A grouping for APN filter definition";

  leaf filter-type {
    type apn-filter-type;
    description "filter type";
  }

  leaf ace-name {
    when "../filter-type = 'apn:acl-filter'";
    type leafref {
      path "/acl:acls/acl:acl/acl:aces/acl:ace/acl:name";
    }
    description "Access Control Entry name.";
  }
}

container apn {
  presence "Enter apn view.";
  description
    "Application-aware Networking.";
  container ipv6 {
    presence "Enter apn-ipv6 view.";
    description
      "Application-aware Networking IPv6.";
    container global {
      description
        "Configure APN6 global config.";
      leaf apn-id-inherit {
        type apn-id-inherit-type;
        description
          "Enable/disable APN ID inherit.";
      }
    }
  }
}
```

```

    container apn-id-templates {
        description
            "List of APN ID templates.";
        list apn-id-template {
            key "name";
            description
                "Configure an APN ID template.";
            leaf name {
                type string {
                    length "1..31";
                    pattern '[^ \?]*';
                }
                description
                    "APN ID template name.";
            }
        }

        container app-info-fields {
            presence "Enter app-info-fields view.";
            description
                "APP information fields.";
            container app-fields {
                description
                    "List of APP fields.";
                list app-field {
                    key "index";
                    unique "name";
                    max-elements "4";
                    description
                        "Configure an APP field.";
                    leaf index {
                        type uint32 {
                            range "1..255";
                        }
                        description
                            "APP field index.";
                    }
                    leaf name {
                        type string {
                            length "1..15";
                            pattern '[^ \?]*';
                        }
                        must "not(../../../../../user-info-fields/user-fields/user-field[n
ame=current()])";
                        mandatory true;
                        description
                            "APP field name.";
                    }
                    leaf length {
                        type uint32 {

```

```

        range "1..32";
    }
    default "16";
    description
        "APP field length.";
    }
}
}
}
container user-info-fields {
    presence "Enter user-info-fields view.";
    description
        "User information fields.";
    container user-fields {
        description
            "List of user fields.";
        list user-field {
            key "index";
            unique "name";
            max-elements "4";
            description
                "Configure an user field.";
            leaf index {
                type uint32 {
                    range "1..255";
                }
                description
                    "User field index.";
            }
            leaf name {
                type string {
                    length "1..15";
                    pattern '^[^ \?]*';
                }
                must "not(../../../../../app-info-fields/app-fields/app-field[name
=current()])";
                mandatory true;
                description
                    "User field name.";
            }
            leaf length {
                type uint32 {
                    range "1..32";
                }
                default "16";
                description
                    "APP field length.";
            }
        }
    }
}

```

```

    }
  }
} ///apn-id-templates

container apn-id-marking {
  presence "Enter user-info-fields view.";
  description
    "Configure apn id marking.";

    container filter {
      uses apn-filter;
      description
        "The filter which is used to indicate the flow to apply
        APN.";
    }

    leaf apn-ipv6-template {
      type leafref {
        path "/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-template/apn:na
me";
      }
      mandatory true;
      description
        "APN IPv6 template.";
    }
}
container app-fields {
  description
    "List of APP fields.";
  list app-field {
    key "name";
    max-elements "4";
    description
      "Configure an APP field.";
    leaf name {
      type leafref {
        path "/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-template[ap
n:name=current()/../../../../apn:apn-ipv6-template]/apn:app-info-fields/apn:app-fiel
ds/apn:app-field/apn:name";
      }
      description
        "APP field name.";
    }
    leaf value {
      type uint32 {
        range "1..4294967295";
      }
      mandatory true;
      description
        "APP field value.";
    }
  }
}

```

```
    }
  }
  container user-fields {
    description
      "List of user fields.";
    list user-field {
      key "name";
      max-elements "4";
      description
        "Configure an user field.";
      leaf name {
        type leafref {
          path "/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-template[ap
n:name=current()/../../../../apn:apn-ipv6-template]/apn:user-info-fields/apn:user-fi
elds/apn:user-field/apn:name";
        }
        description
          "User field name.";
      }
      leaf value {
        type uint32 {
          range "1..4294967295";
        }
        mandatory true;
        description
          "User field value.";
      }
    }
  }
}
} /// apn-id-marking

container mapping-policys {
  description
    "List of mapping policys.";
  list mapping-policy {
    key "color";
    unique "name";
    description
      "Configure a mapping policy.";
    leaf color {
      type uint32 {
        range "0..4294967295";
      }
      description
        "Color of a mapping policy.";
    }
    leaf name {
      type string {
        length "1..31";
        pattern '[^ \?]*';
      }
    }
  }
}
```

```

    }
    mandatory true;
    description
        "Mapping policy name.";
}
leaf description {
    type string {
        length "1..242";
    }
    description
        "Description of a mapping policy.";
}

leaf apn-id-template {
    /// when "../match-type='apn-ipv6'";
    type leafref {
        path "/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-template/
apn:name";
    }
    must "(count(/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-temp
late[apn:name=current()]/apn:app-info-fields/apn:app-fields/apn:app-field) + coun
t(/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-template[apn:name=current()]/
apn:user-info-fields/apn:user-fields/apn:user-field) >= 1)";
    description
        "APN ID template.";
}

container apn-ipv6-maps {
    /// when "../match-type='apn-ipv6'";
    description
        "List of APN IPv6 maps.";
    list apn-ipv6-map {
        key "index";
        description
            "Configure an APN IPv6 map.";
        leaf index {
            type uint32 {
                range "1..4294967295";
            }
            must "((../index = 4294967295 and (count(..app-fields/app-fiel
d) + count(..user-fields/user-field)) = 0) or (../index != 4294967295 and (count
(..app-fields/app-field) + count(..user-fields/user-field)) > 0))";
            description
                "Index.";
        }
        container app-fields {
            when "../index != 4294967295";
            description
                "List of APP fields.";
            list app-field {
                key "name";
                max-elements "4";
                description
                    "Configure an APP field.";
            }
        }
    }
}

```



```
        leaf name {
            type leafref {
                path "/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-t
emplate[apn:name=current()../../../../../apn-id-template]/apn:app-info-fields/ap
n:app-fields/apn:app-field/apn:name";
            }
            description
                "APP field name.";
        }
        leaf value {
            type uint32 {
                range "1..4294967295";
            }
            mandatory true;
            description
                "APP field value.";
        }
    }
}
container user-fields {
    when "../index != 4294967295";
    description
        "List of user fields.";
    list user-field {
        key "name";
        max-elements "4";
        description
            "Configure an user field.";
        leaf name {
            type leafref {
                path "/apn:apn/apn:ipv6/apn:apn-id-templates/apn:apn-id-t
emplate[apn:name=current()../../../../../apn-id-template]/apn:user-info-fields/a
pn:user-fields/apn:user-field/apn:name";
            }
            description
                "User field name.";
        }
        leaf value {
            type uint32 {
                range "1..4294967295";
            }
            mandatory true;
            description
                "User field value.";
        }
    }
}
choice match-tunnel {
    mandatory true;
    description
        "Match tunnel.";
    case sr-policy {
```

```

        description
            "Flow match sr-policy.";
        leaf color {
            type uint32 {
                range "0..4294967295";
            }
            must "not(..../apn-ipv6-map[color=current()][index!=current()../index])";
            description
                "Color of an SR Policy.";
        }
    }
    case ip {
        description
            "Flow match native-ip.";
        leaf native-ip {
            type empty;
            must "not(..../apn-ipv6-map[index!=current()../index]/native-ip)";
            description
                "Native-ip configured.";
        }
    }
}
}
}
}
}
}
} /// mapping-policys
}
}
}
}
}

```

5. IANA Considerations

RFC Ed.: In this section, replace all occurrences of 'XXXX' with the actual RFC number (and remove this note).

IANA is requested to assign a new URI from the IETF XML Registry [RFC3688]. The following URI is suggested:

URI: urn:ietf:params:xml:ns:yang:ietf-apn

Registrant Contact: The IESG.

XML: N/A; the requested URI is an XML namespace.

This document also requests a new YANG module name in the YANG Module Names registry [RFC7950] with the following suggestion:

name: ietf-apn
namespace: urn:ietf:params:xml:ns:yang:ietf-apn
prefix: apn
reference: RFC XXXX

6. Security Considerations

The NETCONF access control model [RFC6536] provides the means to restrict access for particular NETCONF or RESTCONF users to a preconfigured subset of all available NETCONF or RESTCONF protocol operations and content.

There are a number of data nodes defined in this YANG module that are writable/creatable/deletable (i.e., config true, which is the default). These data nodes may be considered sensitive or vulnerable in some network environments. Write operations (e.g., edit-config) to these data nodes without proper protection can have a negative effect on network operations.

7. Acknowledgements

The authors would like to thank the careful reviews and valuable comments from Mengdi Li, Qingyu Guan, Sheng Fang, and Stefano Previdi.

8. Normative References

- [I-D.li-6man-app-aware-ipv6-network]
Li, Z., Peng, S., Li, C., Xie, C., Voyer, D., Li, X., Liu, P., Cao, C., and K. Ebisawa, "Application-aware IPv6 Networking (APN6) Encapsulation", Work in Progress, Internet-Draft, draft-li-6man-app-aware-ipv6-network-03, 22 February 2021, <<https://www.ietf.org/archive/id/draft-li-6man-app-aware-ipv6-network-03.txt>>.
- [I-D.li-apn-framework]
Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C., and G. Mishra, "Application-aware Networking (APN) Framework", Work in Progress, Internet-Draft, draft-li-apn-framework-05, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-framework-05.txt>>.
- [I-D.li-apn-header]
Li, Z., Peng, S., and S. Zhang, "Application-aware Networking (APN) Header", Work in Progress, Internet-

Draft, draft-li-apn-header-02, 7 April 2022,
<<https://www.ietf.org/archive/id/draft-li-apn-header-02.txt>>.

[I-D.li-apn-problem-statement-usecases]

Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z.,
and G. Mishra, "Problem Statement and Use Cases of
Application-aware Networking (APN)", Work in Progress,
Internet-Draft, draft-li-apn-problem-statement-usecases-
06, 7 March 2022, <<https://www.ietf.org/archive/id/draft-li-apn-problem-statement-usecases-06.txt>>.

[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>>.

[RFC6536] Bierman, A. and M. Bjorklund, "Network Configuration
Protocol (NETCONF) Access Control Model", RFC 6536,
DOI 10.17487/RFC6536, March 2012,
<<https://www.rfc-editor.org/info/rfc6536>>.

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

[RFC8340] Bjorklund, M. and L. Berger, Ed., "YANG Tree Diagrams",
BCP 215, RFC 8340, DOI 10.17487/RFC8340, March 2018,
<<https://www.rfc-editor.org/info/rfc8340>>.

[RFC8342] Bjorklund, M., Schoenwaelder, J., Shafer, P., Watsen, K.,
and R. Wilton, "Network Management Datastore Architecture
(NMDA)", RFC 8342, DOI 10.17487/RFC8342, March 2018,
<<https://www.rfc-editor.org/info/rfc8342>>.

[RFC8519] Jethanandani, M., Agarwal, S., Huang, L., and D. Blair,
"YANG Data Model for Network Access Control Lists (ACLs)",
RFC 8519, DOI 10.17487/RFC8519, March 2019,
<<https://www.rfc-editor.org/info/rfc8519>>.

Authors' Addresses

Shuping Peng
Huawei Technologies
Beijing
China
Email: pengshuping@huawei.com

Zhenbin Li
Huawei Technologies
Beijing
China
Email: lizhenbin@huawei.com

RTGWG
Internet-Draft
Intended status: Informational
Expires: 28 April 2022

M. Wang
Q. Cai
L. Han
China Mobile
R. Chen
ZTE Corporation
25 October 2021

cloud-network integration
draft-wang-rtgwg-cloud-network-integration-00

Abstract

This document describes cloud-network integration scenario and networking technologies.

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 28 April 2022.

Copyright Notice

Copyright (c) 2021 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Requirements Language	3
3. Interworking scenarios	3
3.1. Multiple domains with common border nodes	3
3.2. Multiple domains with no common border nodes	4
4. Networking Technologies	4
4.1. Metro network does not support SRv6	5
4.2. Some nodes of the metro network support SRv6	5
4.3. Metro network support SRv6	5
5. Acknowledgements	5
6. IANA Considerations	6
7. Security Considerations	6
8. Normative References	6
Authors' Addresses	6

1. Introduction

With the development of Internet+, the convergence trend of cloud and network is increasingly obvious. More and more services and applications will be carried on the cloud data centers. In order to support new services and applications requirements and meet the security requirements for data not going out of the park, therefore the deployment location of the cloud/data center is also lowered from the original regional DC and core DC to the edge DC.

As the interconnection network between the regional DC and the core DC, the cloud transport network is usually a backbone network. However, with the deployment of the edge DC, in order to avoid new construction of a huge cloud transport network, the existing metro network is used to access the edge DC. The interconnection between edge DCs and regional DC/core DCs is implemented through the coordination between the metro and cloud transport network. Therefore, the interconnection solution between the cloud transport and metro network needs to be considered.

In addition, the access point of enterprises entering the cloud is usually in the metro network, and the dedicated line entering the cloud also involves the interconnection between the cloud transport and metro network.

This document describes cloud-network integration scenario and networking technologies.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

cloud transport network: It is usually a national or province backbone network to achieve interconnection between multiple regional clouds/core clouds deployed in the country/province.

3. Interworking scenarios

This section defines two interworking scenarios.

3.1. Multiple domains with common border nodes

In this scenario, the boundary node of the cloud transport network serves as the boundary node of the metro network. As shown in the figure below. Node 4 serves as the boundary node of the metro network as well as the boundary node of the cloud transport network.

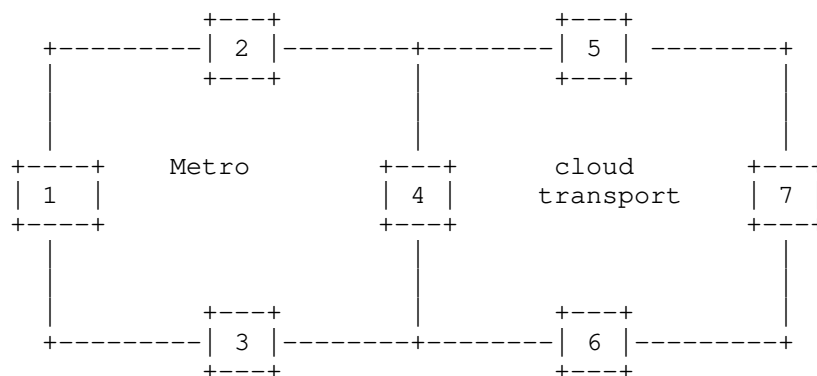


Figure 1

The following applies to the reference topology above:

- * Independent IGP instance in metro region.
- * Independent IGP instance in cloud transport region.
- * If the scale of the metro network is large, sometimes it may reach thousands or even tens of thousands of nodes. At this time, the metro network will be divided into multiple IGPs.

- * The cloud transport and metro network can have different controllers or under the same controller.

3.2. Multiple domains with no common border nodes

In this scenario, the cloud transport network and the metro network do not have a common border nodes, and the border node of the two networks are connected by a direct link. As shown below.

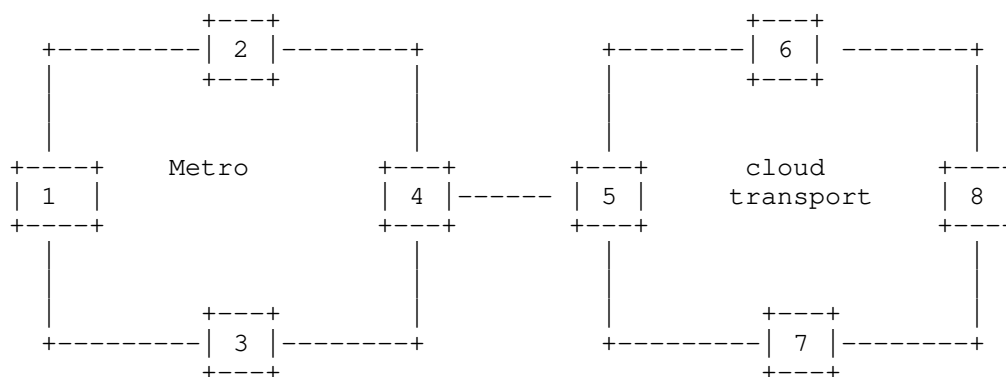


Figure 2

In the interworking scenario described in Section 3.1, since two domains have the same domain boundary node, so the route mutual import can be used by the border node to interconnect the two domains. In this section, the EBGp needs to be deployed between the domains to connect the routes of the two domains.

In this scenario, hierarchical controller architecture usually be considered, that is, the cloud transport and metro network have an independent controller, and cross-domain controllers are used to achieve the coordination of the two domains. If two domains need to be under the same controller, higher requirements are required, such as the controller needs to support a standardized unified southbound interface and so on.

4. Networking Technologies

This section defines three networking technologies.

4.1. Metro network does not support SRv6

Based on existing networks, typically, the metro network does not support the SRv6 and does not have the ability to upgrade to support SRv6. For example, the earlier deployed metro network supports LDP/RSVP/MPLS-TP and traditional L2VPN or L3VPN services. However, the recently deployed metro network may support SR-MPLS/SR-TP, but it still cannot support SRv6 due to its hardware capability.

In this scenario, segment splicing of different network technologies is mainly used to achieve end-to-end connection of services.

4.2. Some nodes of the metro network support SRv6

In some cases, the metro network device connected to the edge DC will be upgraded or replaced to support SRv6, while the rest of the devices should be kept as old as possible and not replaced, so as to avoid the need for more cost investment or avoid affecting the existing services of the metro network.

As shown in Figure 1 or Figure 2, node 4 in metro network is upgraded to support SRv6, while the remaining nodes in metro network do not support SRv6. Cloud transport network supports SRv6. In this scenario, SRv6 is used for end-to-end service connection. The main consideration is how end-to-end SRv6 traverse non-SRv6 networks.

Take figure 1 as an example, the metro network supports SR-MPLS, and Cloud transport network supports SRv6.

[I-D.agrawal-spring-srv6-mpls-interworking] can be used to achieve interworking. In other interworking scenarios, or other metro network scenarios (such as metro networks support LDP/RSVP/MPLS-TP/SR-TP, etc.), the solution needs further discussion.

4.3. Metro network support SRv6

The metro network is a new network that supports SRv6, or a recently deployed network that has the ability to support SRv6 after an upgrade. Therefore, the metro network and cloud transport network are the interworking of two SRv6 domains. In this case, Solutions for interworking between two SRv6 domains need to be considered, including the centralized controller and the distributed control plane solution, and how to implement end-to-end traffic engineering.

5. Acknowledgements

TBD.

6. IANA Considerations

This document makes no request of IANA.

7. Security Considerations

TBD.

8. Normative References

- [I-D.agrawal-spring-srv6-mpls-interworking]
Agrawal, S., ALI, Z., Filsfils, C., Voyer, D., and Z. Li,
"SRv6 and MPLS interworking", Work in Progress, Internet-
Draft, draft-agrawal-spring-srv6-mpls-interworking-06, 22
August 2021, <<https://datatracker.ietf.org/doc/html/draft-agrawal-spring-srv6-mpls-interworking-06>>.
- [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>>.

Authors' Addresses

Minxue Wang
China Mobile
Beijing
China

Email: wangminxue@chinamobile.com

Qian Cai
China Mobile
Beijing
China

Email: caiqian@chinamobile.com

Liuyan Han
China Mobile
Beijing
China

Email: hanliuyan@chinamobile.com

Ran Chen
ZTE Corporation
Nanjing
China

Email: chen.ran@zte.com.cn