

Network Working Group
Internet-Draft
Intended status: Experimental
Expires: 22 June 2022

H. Chen
China Telecom
Z. Hu
Huawei Technologies
H. Chen
Futurewei
X. Geng
Huawei Technologies
Y. Liu
China Mobile
G. Mishra
Verizon Inc.
19 December 2021

SRv6 Midpoint Protection
draft-chen-rtgwg-srv6-midpoint-protection-06

Abstract

The current local repair mechanism, e.g., TI-LFA, allows local repair actions on the direct neighbors of the failed node to temporarily route traffic to the destination. This mechanism could not work properly when the failure happens in the destination point or the link connected to the destination. In SRv6 TE, the IPv6 destination address in the outer IPv6 header could be the dedicated endpoint of the TE path rather than the destination of the TE path. When the endpoint fails, local repair couldn't work on the direct neighbor of the failed endpoint either. This document defines midpoint protection for SRv6 TE path, which enables the direct neighbor of the failed endpoint to do the function of the endpoint, replace the IPv6 destination address to the other endpoint, and choose the next hop based on the new destination address.

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

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 22 June 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 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. SRv6 Midpoint Protection Mechanism	3
3. SRv6 Midpoint Protection Example	3
4. SRv6 Midpoint Protection Behavior	5
4.1. Transit Node as Repair Node	5
4.2. Endpoint Node as Repair Node	6
4.3. Endpoint x Node as Repair Node	6
5. Determining whether the Endpoint could Be Bypassed	7
6. Security Considerations	7
7. IANA Considerations	7
8. Acknowledgements	8
9. References	8
9.1. Normative References	8
9.2. Informative References	8
Authors' Addresses	9

1. Introduction

The current mechanism, e.g., TI-LFA ([I-D.ietf-rtgwg-segment-routing-ti-lfa]), allows local repair actions on the direct neighbors of the failed node to temporarily route traffic to the destination. This mechanism could not work properly when the failure happens in the destination point or the link connected to the destination. In SRv6 TE, the IPv6 destination address in the outer IPv6 header could be the dedicated endpoint of the TE path rather than the destination of the TE path ([RFC8986]). When the endpoint fails, local repair couldn't work on the direct neighbor of the failed endpoint either. This document defines midpoint protection for SRv6 TE path, which enables the direct neighbor of the failed endpoint to do the function of the endpoint, replace the IPv6 destination address to the other endpoint, and choose the next hop based on the new destination address.

2. SRv6 Midpoint Protection Mechanism

When an endpoint node fails, the packet needs to bypass the failed endpoint node and be forwarded to the next endpoint node of the failed endpoint. There are two stages or time periods after an endpoint node fails. The first is the time period from the failure until the IGP converges on the failure. The second is the time period after the IGP converges on the failure.

During the first time period, the packet will be sent to the direct neighbor of the failed endpoint node. After detecting the failure of its interface to the failed endpoint node, the neighbor forwards the packets around the failed endpoint node. It changes the IPv6 destination address with the IPv6 address of the next endpoint node (or the last or other reasonable endpoint node) which could avoid going through the failed endpoint.

During the second time period, the packet of a SRv6 TE path may not be sent to the direct neighbor of the failed endpoint node. There is no route to the failed endpoint node after the IGP converges. When a previous hop node of the failed endpoint node finds out that there is no route to the IPv6 destination address (of the failed endpoint node), it changes the IPv6 destination address with the IPv6 address of the next endpoint node. Note that the previous hop node may not be the direct neighbor of the failed endpoint node.

3. SRv6 Midpoint Protection Example

The topology in Figure 1 illustrates an example of network topology with SRv6 enabled on each node.

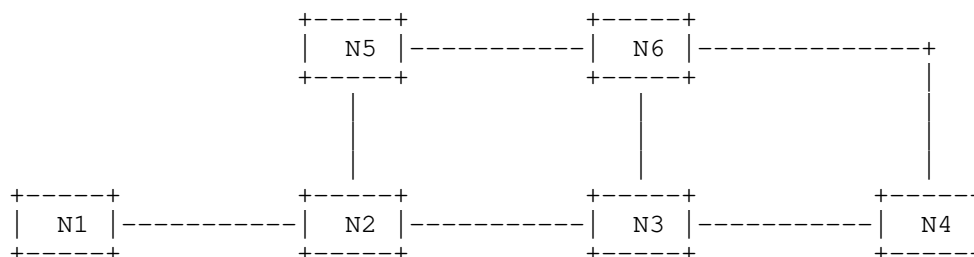


Figure 1: An example of network for midpoint protection

In this document, an end SID at node n with locator block B is represented as $B:n$. An end.x SID at node n towards node k with locator block B is represented as $B:n:k$. A SID list is represented as $\langle S1, S2, S3 \rangle$ where $S1$ is the first SID to visit, $S2$ is the second SID to visit and $S3$ is the last SID to visit along the SRv6 TE path.

In the reference topology, suppose that Node $N1$ is an ingress node of SRv6 TE path going through $N3$ and $N4$. Node $N1$ steers a packet into a segment list $\langle B:3, B:4 \rangle$.

When node $N3$ fails, the packet needs to bypass the failed endpoint node and be forwarded to the next endpoint node after the failed endpoint in the TE path. When outbound interface failure happens in the Repair Node (which is not limited to the previous hop node of the failed endpoint node), it performs the proxy forwarding as follows:

During the first time period (i.e., before the IGP converges), node $N2$ (direct neighbor of $N3$) as a Repair Node forwards the packets around the failed endpoint $N3$ after detecting the failure of the outbound interface to the endpoint $B:3$. It changes the IPv6 destination address with the next sid $B:4$. $N2$ detects the failure of outbound interface to $B:4$ in the current route, it could use the normal Ti-LFA repair path to forward the packet, because it is not directly connected to the node $N4$. $N2$ encapsulates the packet with the segment list $\langle B:5:6 \rangle$ as a repair path.

During the second time period (i.e., after the IGP converges), node $N1$ does not have any route to the failed endpoint $N3$ in its FIB. Node $N1$, as a Repair Node, forwards the packets around the failed endpoint $N3$ to the next endpoint node (e.g., $N4$) directly. There is no need to check whether the failed endpoint node is directly connected to $N1$. $N1$ changes the IPv6 destination address with the next sid $B:4$. Since IGP has completed convergence, it forwards packets directly based on the IGP SPF path

4. SRv6 Midpoint Protection Behavior

A node N protecting the failure of an endpoint node on a SRv6 path may be one of the following types:

- * a transit node: the destination address (DA) of the packet received by N is not N's local SID.
- * an endpoint node: the destination address (DA) of the packet received by N is a N's local END SID.
- * an endpoint x node (i.e., an endpoint with cross-connect node): the destination address (DA) of the packet received by N is a N's local End.X SID with an array of layer 3 adjacencies.

This section describes the behavior of each of these nodes as a repair node for the two time periods after the endpoint node fails.

4.1. Transit Node as Repair Node

When the Repair Node is a transit node, it provides fast protection against the endpoint node failure as follows after looking up the FIB.

```
IF the primary outbound interface used to forward the packet failed
  IF NH = SRH && SL != 0 and
    the failed endpoint is directly connected to Repair Node THEN
    SL decreases*; update the IPv6 DA with SRH[SL];
    FIB lookup on the updated DA;
    forward the packet according to the matched entry;
  ELSE
    forward the packet according to the backup nexthop;
ELSE IF there is no FIB entry for forwarding the packet THEN
  IF NH = SRH && SL != 0 THEN
    SL decreases*; update the IPv6 DA with SRH[SL];
    FIB lookup on the updated DA;
    forward the packet according to the matched entry;
  ELSE
    drop the packet;
ELSE
  forward accordingly to the matched entry;
```

*: SL could be decreased by any dedicated value from [1-N], where N is the current value of SL.

4.2. Endpoint Node as Repair Node

When the Repair Node is an endpoint node, it provides fast protections for the failure through executing the following procedure after looking up the FIB for the updated DA.

```
IF the primary outbound interface used to forward the packet failed
  IF NH = SRH && SL != 0 and
    the failed endpoint is directly connected to Repair Node THEN
    SL decreases; update the IPv6 DA with SRH[SL];
    FIB lookup on the updated DA;
    forward the packet according to the matched entry;
  ELSE
    forward the packet according to the backup nexthop;
ELSE IF there is no FIB entry for forwarding the packet THEN
  IF NH = SRH && SL != 0 THEN
    SL decreases; update the IPv6 DA with SRH[SL];
    FIB lookup on the updated DA;
    forward the packet according to the matched entry;
  ELSE
    drop the packet;
ELSE
  forward accordingly to the matched entry;
```

4.3. Endpoint x Node as Repair Node

When the Repair Node is an endpoint x node, it provides fast protections for the failure through executing the following procedure after updating DA.

```
IF the layer-3 adjacency interface is down THEN
  FIB lookup on the updated DA;
  IF the primary interface used to forward the packet failed THEN
    IF NH = SRH && SL != 0 and
      the failed endpoint directly connected to Repair Node THEN
      SL decreases; update the IPv6 DA with SRH[SL];
      FIB lookup on the updated DA;
      forward the packet according to the matched entry;
    ELSE
      forward the packet according to the backup nexthop;
  ELSE IF there is no FIB entry for forwarding the packet THEN
    IF NH = SRH && SL != 0 THEN
      SL decreases; update the IPv6 DA with SRH[SL];
      FIB lookup on the updated DA;
      forward the packet according to the matched entry;
    ELSE
      drop the packet;
  ELSE
    forward accordingly to the matched entry;
```

5. Determining whether the Endpoint could Be Bypassed

SRv6 Midpoint Protection provides a mechanism to bypass a failed endpoint. But in some scenarios, some important functions may be implemented in the bypassed failed endpoints that should not be bypassed, such as firewall functionality or In-situ Flow Information Telemetry of a specified path. Therefore, a mechanism is needed to indicate whether an endpoint can be bypassed or not. [I-D.li-rtgwg-enhanced-ti-lfa] provides method to determine whether enable SRv6 midpoint protection or not by defining a "no bypass" flag for the SIDs in IGP.

6. Security Considerations

This section reviews security considerations related to SRv6 Midpoint protection processing discussed in this document. To ensure that the Repair node does not modify the SRH header Encapsulated by nodes outside the SRv6 Domain. Only the segment within the SRH is same domain as the repair node. So it is necessary to check the skipped segment have same block as repair node.

7. IANA Considerations

This document makes no request of IANA.

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

8. Acknowledgements

9. References

9.1. 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", Work in Progress, Internet-Draft, draft-ietf-lsr-isis-srv6-extensions-18, 20 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-lsr-isis-srv6-extensions-18.txt>>.
- [I-D.ietf-lsr-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", Work in Progress, Internet-Draft, draft-ietf-lsr-ospfv3-srv6-extensions-03, 19 November 2021, <<https://www.ietf.org/archive/id/draft-ietf-lsr-ospfv3-srv6-extensions-03.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>>.
- [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>>.
- [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>>.
- [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>>.

9.2. 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", Work in Progress, Internet-Draft, draft-hu-spring-segment-routing-

proxy-forwarding-15, 24 October 2021,
<<https://www.ietf.org/archive/id/draft-hu-spring-segment-routing-proxy-forwarding-15.txt>>.

[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", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-07, 29 June 2021, <<https://www.ietf.org/archive/id/draft-ietf-rtgwg-segment-routing-ti-lfa-07.txt>>.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", Work in Progress, Internet-Draft, draft-ietf-spring-segment-routing-policy-14, 25 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-spring-segment-routing-policy-14.txt>>.

[I-D.li-rtgwg-enhanced-ti-lfa]
Li, C., Hu, Z., Zhu, Y., and S. Hegde, "Enhanced Topology Independent Loop-free Alternate Fast Re-route", Work in Progress, Internet-Draft, draft-li-rtgwg-enhanced-ti-lfa-05, 21 October 2021, <<https://www.ietf.org/archive/id/draft-li-rtgwg-enhanced-ti-lfa-05.txt>>.

[I-D.sivabalan-pce-binding-label-sid]
Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J., Previdi, S., and C. Li, "Carrying Binding Label/Segment-ID in PCE-based Networks.", Work in Progress, Internet-Draft, draft-sivabalan-pce-binding-label-sid-07, 8 July 2019, <<https://www.ietf.org/archive/id/draft-sivabalan-pce-binding-label-sid-07.txt>>.

[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

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

Email: chenhuan6@chinatelecom.cn

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

Email: huzhibo@huawei.com

Huaimo Chen
Futurewei
Boston, MA,
United States of America

Email: Huaimo.chen@futurewei.com

Xuesong Geng
Huawei Technologies

Email: gengxuesong@huawei.com

Yisong Liu
China Mobile

Email: liuyisong@chinamobile.com

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring, MD 20904
United States of America

Phone: 301 502-1347
Email: gyan.s.mishra@verizon.com

RTGWG
Internet-Draft
Intended status: Informational
Expires: 27 October 2022

D. King
Lancaster University
A. Farrel
Old Dog Consulting
C. Jacquenet
Orange
25 April 2022

Challenges for the Internet Routing Systems Introduced by Semantic
Routing
draft-king-irtf-challenges-in-routing-08

Abstract

Historically, the meaning of an IP address has been to identify an interface on a network device. Routing protocols were developed based on the assumption that a destination address had this semantic.

Over time, routing decisions have been enhanced to determine paths on which packets could be forwarded according to additional information carried principally within the packet headers, and dependent on policy coded in, configured at, or signaled to the routers.

Many proposals have been made to add semantics to IP packets by placing additional information into existing fields, by adding semantics to IP addresses, or by adding fields to the packets. The intent is always to facilitate routing decisions based on these additional semantics in order to provide differentiated paths to enable forwarding of different packet flows on paths that may be distinct from those derived by shortest path first or path vector routing. We call this approach "Semantic Routing".

This document describes the challenges to the existing routing system that are introduced by Semantic Routing. It then summarizes the opportunities for research into new or modified routing and forwarding approaches that make use of additional semantics.

This document is presented as a study to support further research into clarifying and understanding the issues. It does not pass comment on the advisability or practicality of any of the proposals and does not define any technical solutions.

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 27 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. Current Challenges to IP Routing	4
3. What is Semantic Routing?	7
3.1. Architectural Considerations	9
4. Challenges for Internet Routing Research	10
4.1. Research Principles	10
4.2. Routing Research Questions to be Addressed	11
5. Security and Privacy Considerations	15
6. IANA Considerations	16
7. Acknowledgements	16
8. Contributors	16
9. Informative References	16
Authors' Addresses	17

1. Introduction

Historically, the meaning of an IP address has been to identify an interface on a network device. Routing protocols were to compute, establish, and maintain paths through networks toward destination prefixes until IP packets eventually reach their destination, and were based on the assumption that a destination address had this semantic. Anycast and multicast addresses were also defined, and those address semantics sometimes required variations to the routing protocols or even encouraged the development of new protocols.

Over time, the mechanisms that enabled routing decisions were enhanced to determine paths on which packets could be forwarded according to additional information carried principally within the packets headers or within 'shim' headers, and dependent on policy coded in, configured at, or signaled to the routers. Perhaps one of the most iconic examples is Equal-Cost Multipath (ECMP) where a router makes a choice about how to forward a packet over a number of parallel links or paths based on the values of a set of fields in the packet header.

Many proposals have been made to add semantics to IP packets by placing additional information into existing fields, by adding semantics to IP addresses, or by adding fields to the packets. The intent is always to facilitate routing decisions based on these additional semantics in order to provide differentiated paths to enable forwarding of different packet flows on paths that may be distinct from those derived by shortest path first or path vector routing. We call this approach "Semantic Routing"
[I-D.farrel-irtf-introduction-to-semantic-routing].

There are many approaches to adding semantics to packet headers: the additional information may be derived from the destination addresses, from other fields in the packet header, or the packet itself. Mechanisms for using the destination address range from assigning an address prefix to have a special purpose and meaning (such as is done for multicast addressing) through allowing the owner of a prefix to use the low-order bits of an address for specific purposes (e.g., to provide an indication of the nature of the service that is associated with these packets). Some proposals suggest variable address lengths, others offer new hierarchical address formats, and some introduce a structure to addresses so that they can carry additional information in a common way. Alternatively, forwarding decisions can be performed based on fields in the packet header (such as the IPv6 Flow Label, or the Traffic Class field), overloading of existing packet fields, or new fields added to the packet headers.

A survey of ways in which routing and forwarding decisions have been made based on additional information carried in packets can be found in [I-D.king-irtf-semantic-routing-survey].

Some Semantic Routing proposals are intended to be deployed in administratively scoped IP domains whose network components (routers, switches, etc.) are operated by a single administrative entity (sometimes referred to as 'limited domains' [RFC8799]), while other proposals are intended for use across the Internet. The impact the proposals have on routing systems may require clean-slate solutions, hybrid solutions, extensions to existing routing protocols, or potentially no changes at all.

This document describes some of the key challenges to the routing system that are already present in today's IP networks. It then briefly outlines the concept of "Semantic Routing" with reference to [I-D.farrel-irtf-introduction-to-semantic-routing] and presents some of the additional challenges to the existing routing system that Semantic Routing may introduce. Finally, this document presents a list of research questions that offer opportunities for future research into new or modified routing protocols and forwarding systems that make use of Semantic Routing.

In this document, the focus is on routing and forwarding at the IP layer. A variety of overlay mechanisms exists to perform service or path routing at higher layers, and those approaches may be based on similar extensions to packet semantics, but that is out of scope for this document. Similarly, it is possible that Semantic Routing can be applied in a number of underlay network technologies, and that, too, is out of scope for this document.

This document is presented as a study to support further research into clarifying and understanding the issues. It does not pass comment on the advisability or practicality of any of the proposals and does not define any technical solutions.

2. Current Challenges to IP Routing

Today's IP routing faces several significant challenges which are a consequence of architectural design decisions and the continued exponential growth in traffic. These challenges include mobility, multihoming, programmable paths, scalability, and security, and were not the focus of the original design of the Internet. Nevertheless, IP networks have, in general, coped well in an incremental manner whenever a new challenge has arisen. The following list is presented to give context to the continuing requirements that routing protocols must meet as new semantics are applied to the routing process.

- * Mobility - Mobility introduces several challenges, including maintaining a relationship between a sender and a receiver in cases where the sender or receiver changes their point of network attachment. The network must always be informed about the mobile node's current location, to allow continuity of services. Mobile users may also consume network resources, while in motion. The mobile user's service instances and attachments will also change due to varying load or latency, e.g., in Multi-access Edge Computing (MEC) environments.
- * Multihoming - Multihomed stations or multihomed networks are connected to the Internet via more than one access circuit or access network and, therefore, may be assigned multiple IP addresses or prefixes from different pools. There are challenges concerning how traffic is forwarded back to the source if the source has originated its traffic using the wrong source address for a particular connection, or if one of the connections to the Internet is degraded.
- * Multi-path - The Internet was initially designed to find the single, "best" path to a destination using a distributed routing algorithm. Current IP network topologies can provide multiple paths to reach a destination, each with different characteristics and with different failure likelihoods. It may be beneficial to send traffic over multiple paths to achieve reliability and enhance throughput, and it may be desirable to select one path or another because of QoS or security considerations for example, or to avoid transiting specific areas of an IP network, based (for example) on the reputation of transit provider for example. However, how packets are forwarded by using the shortest path means that distinguishing these alternate paths and directing traffic to them can be hard. Further, problems concerning scalability, commercial agreements among Service Providers, and the design of BGP make the utilization of multi-path techniques difficult for inter-domain routing. (Note that this discussion is distinct from Equal Cost Multi-path (ECMP) where packets are directed onto several "parallel" paths of identical least cost using a hash algorithm operated on some of the packets' header fields.)
- * Multicast - Delivering the same packet to multiple destinations can place considerable load on a network. Solutions that replicate the packet at the source or at the network edge may obviously cause multiple copies of the packet to flow along the same network links. Solutions that move deterministic replication into the network to make more optimal use of the network resources can be complex to set up and manage since multicast network designs often assume dynamic tree computation where the multicast

distribution tree can be rooted at the source or in the multicast network, thereby leading to specific routing tables whose entries denote the tree structure. More complicated hardware that can replicate packets may also be required within the network. In order that packets can be addressed to a group of destinations and not be forwarded by means of unicast transmission, parts of the addressing space (that is, address prefixes) have been reserved for multicast addressing.

- * Programmable Paths - The ability to decouple IP paths from routing protocols and agreements between Service Providers could allow users and applications to select network paths themselves, based on the required path characteristics. Another option is to let the route computation logic select, establish, and maintain paths on behalf of the user or the application and as a function of their requirements so that Service Providers can participate in the route computation "service". Currently, user and application packets follow the path selected by routing protocols and the way traffic is forwarded through a network is under the control of the Service Provider that operates the said network. The corresponding traffic forwarding policies enforced by the service provider usually comply with the requirements expressed by the user or the application. These requirements may have triggered a dynamic service parameter negotiation cycle that eventually leads to proper (network, CPU, storage) resource allocation.
- * Endpoint Selection - As compute resources and content storage move closer to the edge of the network, there are often multiple points in the network that can satisfy user requests. In order to make the best use of these distributed resources and so as to not overload parts of the network, user traffic needs to be steered to appropriate servers or data centres. In many cases, this function may be achieved in the application layer (such as through DNS [RFC3467]) or in the transport layer (such as using ALTO [RFC5693]). The challenge is to balance higher-layer decisions about which application layer resources to use with information from the lower layers about the availability and load of network resources.
- * Scalability - There are many scaling concerns that pose critical challenges to the Internet. Not least among these challenges is the size of the routing tables that routers in an IP network must maintain. As the number of devices attached to the network grows, so the number of addresses in use also grows, and because of the schemes used to assign address prefixes, the mobility of devices, and the various connectivity options between networks, the routing table sizes also grow, even more so when prefixes are not always amenable to aggregation. This problem is exacerbated by some

services (such as those supported by the IoT where several thousands of objects/sensors may be networked), where, as more devices are added to the network, the size of the routing table may affect the operation of certain routing protocols. It may be noted that scaling issues are also exacerbated by multihoming practices if a host that is multihomed is allocated a different address for each point of attachment.

- * Manageability, Maintainability, and Extensibility - Operational manageability is a key requirement for network technologies: network operators must be able to determine the status of their network and understand the causes of any disruptions or problems. Further, it must be possible to maintain the networks and the technologies running in them without disrupting the services being delivered by the networks. Additionally, the network technologies developed and deployed need to be extensible so that new features can be added and new services supported without the need to invent whole new technologies.
- * Security - Issues of security and privacy have been largely overlooked by the routing systems. However, there is increasing concern that attacks on routing systems can not only be disruptive (for example, causing traffic to be dropped), but may cause traffic to be redirected to inspection points that can breach the security or privacy of the payloads.

Some of the challenges outlined here were previously considered within the IETF by the IAB's "Routing and Addressing Workshop" held in Amsterdam, The Netherlands on October 18-19, 2006 [RFC4984]. Several architectures and protocols have since been developed and worked on within and outside the IETF, and these are examined in [I-D.king-irtf-semantic-routing-survey].

3. What is Semantic Routing?

Semantic Routing is the term applied to routing in an IP network that relies upon additional information to feed the route computation process, to enhance route selection decisions, and to direct the forwarding process. In addition to the routable part of the destination IP address (the prefix), such information may be present in other fields in the packet (chiefly the packet header) and configured or programmed into the routers/forwarders. Semantic Routing includes mechanisms such as "Preferential Routing", "Policy-based Routing", and "Flow steering".

In Semantic Routing, a packet forwarding engine may examine a variety of fields in a packet and match them against forwarding instructions. Those forwarding instructions may be installed by routing protocols,

configured through management protocols or a software defined networking (SDN) controller, or derived by a software component on the router that considers network conditions and traffic loads. The packet fields concerned may be the fields of an IP header, those same fields but with additional semantics, elements of the packet payload, or new fields defined for inclusion in the packet header or as a "shim" between the header and payload. In the case of additional semantics included in existing packet header fields, the approach implies some "overloading" of those fields to include meaning beyond the original definition. In all cases, a well-known definition of the encoding of the additional information is required to enable consistent interpretation within the network.

A more detailed description of Semantic routing can be found in [I-D.farrel-irtf-introduction-to-semantic-routing] and a survey of Semantic Routing proposals and research projects can be found in [I-D.king-irtf-semantic-routing-survey].

Many technical challenges exist for Semantic Routing in IP networks depending on which approach is taken. These challenges include (but are not limited to):

- * The continual growth of routing tables.
- * Convergence times for large networks.
- * Granularity of routing decisions.
- * Address consumption caused by lower address utility rate. The wastage mainly comes from aligning finite allocation for semantic address blocks.
- * Encoding too many semantics into prefixes will require evaluation of which to prioritize.
- * Risk of privacy/information leakage.
- * Lack of visibility of the Semantic Routing information when end-to-end or edge-to-edge encryption is used.
- * Burdening the user, application, or prefix assignment node.
- * Source address spoofing prevention mechanisms are required.
- * Overloading of routing protocols causing stability and scaling problems.

- * Depending on encoding mechanisms, there may be challenges for data planes to scale the processes of finding, reading, and looking up semantic data in order to forward packets at line speed.
- * Backwards compatibility with existing IP networking and routing protocols.
- * Extensibility to support additional functions in the future.
- * Manageability and network diagnostics to be able to determine how the network is functioning and to isolate the causes of any problems.

3.1. Architectural Considerations

Semantic data may be taken into account to integrate with existing routing architectures. An overlay can be built such that Semantic Routing is used to forward traffic between nodes in the overlay, but regular IP is used in the underlay. The application of semantics may also be constrained to within a limited domain. In some cases, such a domain will use IP, but be disconnected from the Internet. In other cases, traffic from within the domain is exchanged with other domains that are connected together across an IP network using tunnels or via application gateways. And in still another case traffic from the domain is forwarded across the Internet to other nodes and this requires backward-compatible routing approaches.

Isolated Domains: Some IP network domains are entirely isolated from the Internet and other IP networks. In these cases, packets cannot "escape" from the isolated domain into external networks and so the Semantic Routing schemes applied within the domain can have no detrimental effects on external domains. Thus, the challenges are limited to enabling the desired function within the domain.

Bridged Domains: In some deployments, it will be desirable to connect together multiple isolated domains to build a larger network. These domains may be connected (or bridged) over an IP network or even over the Internet, possibly using tunnels. An alternative to tunneling is achieved using gateway functionality where packets from a domain are mapped at the domain boundary to produce regular IP packets that are sent across the IP network.

Semantic Prefix Domains: A semantic prefix domain is a portion of the Internet over which a consistent set of semantic-based policies are administered in a coordinated fashion. This is achieved by assigning a routable address prefix (or a set of prefixes) for use with Semantic Routing so that packets may be

forwarded through the regular IP network (or the Internet). Once delivered to the semantic prefix domain, a packet can be subjected to whatever Semantic Routing is enabled in the domain.

Further discussion of architectures for Semantic Routing can be found in [I-D.farrel-irtf-introduction-to-semantic-routing].

4. Challenges for Internet Routing Research

It may not be possible to embrace all emerging scenarios with a single approach or solution. Requirements such as 5G mobility, near-space-networking, and networking for outer-space (inter-planetary networking), may need to be handled using different network technologies. Improving IP network capabilities and capacity to scale, and address a set of growing requirements presents significant research challenges, and will require contributions from the networking research community. Solutions need to be both economically feasible and have the support of the networking equipment vendors as well as the network operators.

4.1. Research Principles

Research into Semantic Routing should be founded on regular scientific research principles [royalsoc]. Given the importance of the Internet today, it is critical that research is targeted, rigorous, and reproducible.

The most valuable research will go beyond an initial hypothesis, a report of the work done, and the results observed. Although that is a required foundation, networking research needs to be independently reproducible so that claims can be verified or falsified. Further, the networks on which the research is carried out need to both reflect the characteristics that are being explicitly tested, and reproduce the variety of real networks that constitute the Internet.

Thus, when conducting experiments and research to address the questions in Section 4.2, attention should be given to how the work is documented and how meaningful the test environment is, with a strong emphasis on making it possible for others to reproduce and validate the work.

4.2. Routing Research Questions to be Addressed

As research into the scenarios and possible uses of Semantic Routing progresses, a number of questions need to be answered. These questions go beyond "Why do we need this function?" and "What could we achieve by carrying additional semantics in an IP address?" The questions are also distinct from issues of how the additional semantics can be encoded within an IP address. All of those issues are, of course, important considerations in the debate about Semantic Routing, but they form only part of the essential groundwork of research into Semantic Routing itself.

This section sets out some of the concerns about how the wider the use of Semantic Routing might impact a routing system. These questions need to be answered in separate research work or folded into the discussion of each Semantic Routing proposal.

1. What is the scope of the Semantic Routing proposal? This question may lead to various answers:

Global: It is intended to apply to all uses of IP.

Backbone: It is intended to apply to IP network connectivity.

Overlay: It is to be used as an overlay network using tunneling over IP or other underlay technologies.

Gateway: The Semantic Routing will be used within a specific domain, and communications with the wider Internet will be handled by IP and probably application gateways.

Domain: The use of the Semantic Routing is strictly limited to within a domain or private network.

Underlying this question is a broader question about the boundaries of the use of IP, and the limit of "the Internet". If a limited domain is used, is it a semantic prefix domain [RFC8799] where a part of the IP address space identifies the domain so that an address is routable to the domain, but the additional semantics are used only within the domain, or is the address used exclusively within the domain so that the external impact of the routability of the address and the additional semantics is not important?

2. What will be the impact on existing routing systems? What would happen if a packet carrying additional semantics was subjected to normal routing operations? How would the existing routing systems react if such a packet escaped (accidentally or

maliciously) from the planned scope of the proposal? For example: how are the semantic parts of an address distinguished from the routable parts (if, indeed, they are separable)?; is there an impact on the size and maintenance of routing tables due to the addition of semantics?; how are cryptographically generated addresses (such as [RFC3972]) made routable and kept simple enough for management?.

3. What path characteristics are needed to describe the desired paths and as input to route computation? Since one of the implications of adding semantics to IP packets is to cause special processing by routers, it is important to understand what behaviors are wanted. Such path characteristics include (but are not limited to):

Quality: Expressed in terms of throughput, latency, jitter, drop precedence, etc.

Resilience: Expressed in terms of survival of network failures and delivery guarantees.

Destination: How is a destination address to be interpreted if it encodes a choice of actual destinations? Can traffic be forwarded over multiple distinct paths if multiple destination addresses are encoded?

Security: What choices of path reduce the vulnerability of the traffic to security or privacy attacks?

In these cases, how do the routers utilize the additional semantics to determine the desired characteristics? Or are such characteristics used to feed the route computation logic, for example, by means of metrics? What additional information about the network do the routing protocols need to gather? What changes to the routing algorithm are needed to deliver packets according to the desired characteristics? How can routes be computed with characteristics that accommodate traffic patterns, requirements, and constraints?

4. Can we solve these routing challenges with existing routing tools and methods? We can break this question into a set of more detailed questions.

- * Is new hardware needed? Existing deployed hardware has certain assumptions about how forwarding is carried out based on IP addresses and routing tables. But hardware is increasingly programmable so that it may be possible to instruct the forwarding components to act on a variety of elements of the packets.
- * Do we need new routing protocols? We might ask some subsidiary questions:
 - Can we make do with existing protocols, possibly by tuning configuration parameters or using them out of the box?
 - Can we make backwards-compatible modifications to existing protocols such that they work equally for today's IP addresses or addresses with extra semantics?
 - Do we need entirely new protocols or radical evolutions of existing protocols in order to enforce advanced Semantic Routing policies?
 - Should we focus on the benefits of routing solutions that are optimized for specific environments (network topologies, technologies, use cases), or should we attempt to generalize to enable wider applicability?
- 5. Do we need new management tools and techniques? How practical is it to debug and operate the routing system? Management of the routing system (especially diagnostic management) is a crucial and often neglected part of the problem space. A critical part of this issue is how packets within the network can be inspected by diagnostic tools (or human operators) and mapped to the routing and forwarding decisions that were made within the network in order to understand the actions made at and by upstream routers.
- 6. What is the impact of Semantic Routing on the security of the routing system?
 - * Does the introduction of Semantic Routing provide a greater attack surface?
 - * Can Semantic Routing provide greater opportunities for security by fine-grain forwarding of flows to be inspected by different security functions?

- * Can Semantic Routing improve security and privacy by obscuring information in the packets, or does the inclusion of additional information risk compromising security and privacy?
 - * To what extent does deployment within a limited domain strengthen security or make it less of a concern?
 - * Does the use of Semantic Routing make it easier or harder to impose censorship, prohibit access to the Internet by specific parties, or block access to certain resources or types of service?
7. What is the scalability impact of Semantic Routing on routing systems? Scalability can be measured as:
- * Routing table size. How many entries need to be maintained in the routing tables by different routers serving different roles in the network? Some approaches to Semantic Routing may be explicitly intended to address this problem.
 - * Forwarding table size. The size of the forwarding table may be less of an issue considering modern hardware, however the more granular the routing/forwarding decisions made in a router, the greater the size of this table. The size of the forwarding table has implications for memory in the forwarding engine, but also for the lookup time for forwarding each packet.
 - * Routing performance. Routing performance may be considered in terms of the volume of data that has to be exchanged both to construct and maintain the routing tables at the participating routers. It may also be measured in terms of how much processing is required to compute new routes when there is a change in the network.
 - * Routing convergence. This is the time that it takes for a routing protocol to discover changes (especially faults) in the network, to distribute the information about any changes to its peers, and to reach a stable state across the network such that packets are forwarded consistently.

For all questions about routing scalability, research that presents figures based on credible example networks is highly desirable. Similar questions may be asked about the amount of forwarding state that has to be maintained in the routers.

8. To what extent can Semantic Routing be applied to multicast transmission schemes:
 - * Can Semantic Routing facilitate the computation and the establishment of (service-inferred) multicast distribution trees?
 - * Can specific semantics be carried in multicast addresses?
9. Is the approach extensible and maintainable? Can new features be added without increasing the complexity and in a backward compatible way? Could the approach be modified to handle evolutions in the rest of the networking infrastructure? Considerations might include the ability to encode additional options or variants within protocol fields, and the ability to add new fields. Such considerations must be actively traded against the processing overhead associated with certain encoding types.
10. What aspects need to be standardized? It is important to understand the necessity of standardization within this research. What degree of interoperability is expected between devices and networks? Is a given domain so constrained (for example, to a single equipment vendor) that standardization would be meaningless? Is the application so narrow (for example, in niche hardware environments) such that interoperability is best handled by agreements among small groups of vendors such as in industry consortia?

5. Security and Privacy Considerations

Research into Semantic Routing must give full consideration to the security and privacy issues that are introduced by these mechanisms. Placing additional information into packet header fields might reveal details of what the packet is for, what function the user is performing, who the user is, etc. Furthermore, in-flight modification of the additional information might not directly change the destination of the packet, but might change how the packet is handled within the network and at the destination.

It should also be considered how packet encryption techniques that are increasingly popular for end-to-end or edge-to-edge security may obscure the semantic information carried in some fields of the packet header or found deeper in the packet. This may render some semantic routing techniques impractical and may dictate other methods of carrying the necessary information to enable Semantic Routing.

6. IANA Considerations

This document makes no requests for IANA action.

7. Acknowledgements

Thanks to Stewart Bryant for useful conversations. Luigi Iannone, Robert Raszuk, Dirk Trossen, Ron Bonica, Marie-Jose Montpetit, Yizhou Li, Toerless Eckert, Tony Li, Joel Halpern, Stephen Farrell, Carsten Bormann, David Hutchison, Jeffery He, Dino Farinacci, Greg Mirsky, and Jeff Haas made helpful suggestions.

This work is partially supported by the European Commission under Horizon 2020 grant agreement number 101015857 Secured autonomic traffic management for a Tera of SDN flows (Teraflow).

8. Contributors

Joanna Dang
Email: dangjuanna@huawei.com

9. Informative References

- [I-D.farrel-irtf-introduction-to-semantic-routing]
Farrel, A. and D. King, "An Introduction to Semantic Routing", Work in Progress, Internet-Draft, draft-farrel-irtf-introduction-to-semantic-routing-03, 22 January 2022, <<https://www.ietf.org/archive/id/draft-farrel-irtf-introduction-to-semantic-routing-03.txt>>.
- [I-D.king-irtf-semantic-routing-survey]
King, D. and A. Farrel, "A Survey of Semantic Internet Routing Techniques", Work in Progress, Internet-Draft, draft-king-irtf-semantic-routing-survey-03, 26 November 2021, <<https://www.ietf.org/archive/id/draft-king-irtf-semantic-routing-survey-03.txt>>.
- [RFC3467] Klensin, J., "Role of the Domain Name System (DNS)", RFC 3467, DOI 10.17487/RFC3467, February 2003, <<https://www.rfc-editor.org/info/rfc3467>>.
- [RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, DOI 10.17487/RFC3972, March 2005, <<https://www.rfc-editor.org/info/rfc3972>>.

- [RFC4984] Meyer, D., Ed., Zhang, L., Ed., and K. Fall, Ed., "Report from the IAB Workshop on Routing and Addressing", RFC 4984, DOI 10.17487/RFC4984, September 2007, <<https://www.rfc-editor.org/info/rfc4984>>.
- [RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, DOI 10.17487/RFC5693, October 2009, <<https://www.rfc-editor.org/info/rfc5693>>.
- [RFC8799] Carpenter, B. and B. Liu, "Limited Domains and Internet Protocols", RFC 8799, DOI 10.17487/RFC8799, July 2020, <<https://www.rfc-editor.org/info/rfc8799>>.
- [royalsoc] The Royal Society, "Evidence synthesis : Principles", Web page, Principles for good evidence synthesis, 19 September 2018, <<https://royalsociety.org/topics-policy/projects/evidence-synthesis/principles/>>.

Authors' Addresses

Daniel King
Lancaster University
United Kingdom
Email: d.king@lancaster.ac.uk

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

Christian Jacquenet
Orange
Rennes
France
Email: christian.jacquenet@orange.com

SPRING Working Group
Internet-Draft
Intended status: Informational
Expires: October 2, 2021

C. Li
Z. Li
H. Yang
Huawei Technologies
March 31, 2021

IPv6-based Cloud-Oriented Networking (CON)
draft-li-rtgwg-ipv6-based-con-01

Abstract

This document describes the scenarios, requirements and technologies for IPv6-based Cloud-oriented Networking.

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 2, 2021.

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. Terminology	3
2.1. Requirements Language	4
3. Problem Statement	4
3.1. Underlay	4
3.2. Overlay	6
4. IPv6-based Cloud-Oriented Networking	6
4.1. Requirements	7
4.1.1. Quick Connection	7
4.1.2. Hybrid Network Connection	7
4.1.3. Path Programming	7
4.1.4. Resource Assurance	8
4.1.5. Deterministic Delay	8
4.1.6. Service Function Chaining	8
4.1.7. Performance Measurement	9
4.1.8. Reliability	9
4.1.9. Security	9
4.1.10. Forwarding Efficiency	9
4.1.11. Application-Aware Networking	10
4.2. Solutions	10
4.2.1. VPN	10
4.2.2. Path Programming	11
4.2.3. Service Function Chaining	11
4.2.4. IPv6 based Network Slicing	11
4.2.5. IPv6-based On-path Measurement	12
4.2.6. Reliability	13
4.2.7. Security	14
4.2.8. IPv6 Forwarding Efficiency	14
4.2.9. Application-aware IPv6 Networking	15
5. IANA Considerations	15
6. Security Considerations	15
7. Contributors	16
8. Acknowledgements	16
9. References	16
9.1. Normative References	16
9.2. Informative References	16
Authors' Addresses	21

1. Introduction

With the development of cloud computing, increasing services have been migrated from enterprises to cloud data centers. Compared with interconnections between branches and headquarters, new connections between enterprise sites to cloud data centers and inter-cloud are added, which bring new requirements and challenges for existing networks.

When enterprises have workloads & applications & data split among different data centers, especially for those enterprises with multiple sites that are already interconnected by VPNs (e.g., MPLS L2VPN/L3VPN), challenges are introduced. [I-D.ietf-rtgwg-net2cloud-problem-statement] describes the problems that enterprises face today when interconnecting their branch offices with dynamic workloads in third party data centers (a.k.a. Cloud DCs).

SD-WAN is a flexible WAN architecture that enables flexible network-to-cloud and inter-clouds connections. It supports to use alternative paths like internet or 4G / 5G connection instead of expensive MPLS leased lines to exchange data between sites and clouds. However, when a WAN path travels multiple MPLS domains, the configurations are complicated due to multiple services touch points need to be configured. Therefore, it is hard to support end-to-end path programming in IPv4/MPLS based SD-WAN.

When using VXLAN in SD-WAN, only the overlay path or anchor points can be specified while the underlay forwarding path can not be specified. Therefore, strict TE requirements like deterministic delay, specified path forwarding can not be satisfied.

In order to resolve these challenges, this document propose IPv6-based Cloud-Oriented Networking (CON). In addition, it describes the challenges for existing networks when clouds and networks are converged, requirements that IPv6-based CON should satisfy, and the solutions in IPv6-based CON that satisfy the requirements.

IPv6-based CON supports quick and flexible connections between sites and clouds and inter-clouds, it also supports end-to-end path programming, which can be used for many use cases, such as strict path traffic engineering, deterministic delay forwarding, and service function chaining, to provide better network services for cloud-network and inter-cloud interconnections.

2. Terminology

This document makes use of the terms defined in [RFC8754] and [RFC8200], and the reader is assumed to be familiar with that terminology. This document introduces the following terms:

POP: Point of Presence

CON: Cloud-Oriented Networking.

EC: Edge Computing.

EDC: Edge Data center

RDC: Regional Data Center

CDC: Core Data Center

2.1. 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.

3. Problem Statement

As development of cloud, many clouds have been deployed, such as Private cloud, Public Cloud, and Hybrid Cloud. The cloud services can be provided by a third party, such as an OTT (Over-The-Top) content provider, and it can be provided by a network operator as well. Furthermore, cloud can be deployed not only in IT data centers but also CT data centers.

With the development and successful application of cloud native design in the IT field and Network Functions Virtualization (NFV) technologies, virtualization and cloudification have gradually matured and evolved to provide a new level of productivity, offering a new approach to telecom network construction. Building cloud-based telecom networks (also known as telco clouds) becomes a new way of telecom network construction.

In order to support low latency communication, the request should be responded at the near cloud data center, therefore edge computing data center (a.k.a Edge Cloud) is introduced. Telecommunication services and third-party OTT services can be deployed in the edge cloud.

As the deployment of clouds, the traffic pattern in the network has changed significantly, which results in new challenges for existing networks.

3.1. Underlay

From the aspect of underlay, cloud services requires the network to provide quick and flexible connection.

The typical topology of telco cloud is shown in figure 1.

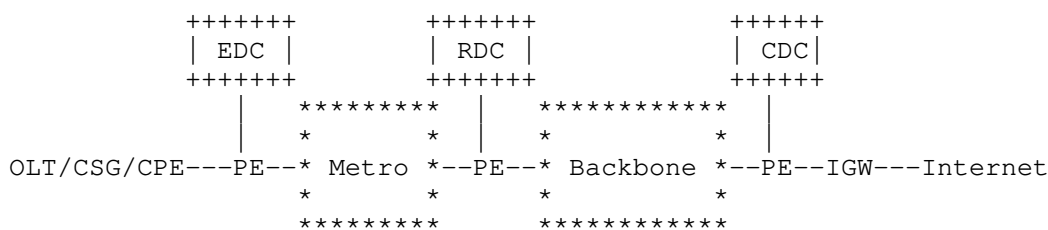


Figure 1. Typical topology of telco cloud

The edge cloud is deployed in edge data center (EDC) in the access network usually, so that the servers in the edge cloud can respond to the delay-sensitive requests rapidly, like 5G URLLC traffic. The traffic is not delay-sensitive can be responded in regional cloud DC, which is located in the metro or core network. Most cloud services may be deployed in the core cloud DC considering reducing the cost, which is far from the end user. Like, the UPF may locate in the regional cloud DC or Core cloud DC, so that it can support more users.

The traffic from end users to cloud servers are forwarded along with different paths due to the different locations of the end users and the cloud servers.

In addition, when deploying new services, for instance, deploying a leased line from an enterprise site to a cloud data center, it will take probably weeks in the current IPv4/MPLS carrier network. Because the VPN configuration is needed to be done at multiple PE nodes if the leased line travels multiple domains (when using Option A between domains). Also, the cloud operator needs to negotiate with network operators if the cloud services and the network services are provided by different operators. For example, thousands of chain stores such as grocery stores or super markets are needed to connect to their enterprise VPNs, and they may use the cloud services. However, in IPv4/MPLS network, it may take weeks to establish a new VPN connection from a site to headquarter or cloud tenant networks.

Furthermore, different traffic of different enterprise/tenant/users are treated differently in clouds, and they MAY be forwarded along with different service function chains (SFC). However, it is hard to support SFC in IPv4 or MPLS based network in carrier's networks or data center networks. Normally, to support SFC, the traffic steering policies are configured at multiple nodes along the SFC path, which is complicated.

3.2. Overlay

In order to provide quick and flexible cloud connection, overlay connection is provided by cloud providers, especially the OTT cloud providers.

SD-WAN is a typical fabric for DCI connection between clouds and sites, which provides a cheaper and smarter WAN connection. Many SD-WAN providers build their own WAN backbone network by connecting their POP GWs to provide better SLA assurance for tenants. The typical topology of SD-WAN with POP GWs is shown in figure 2.

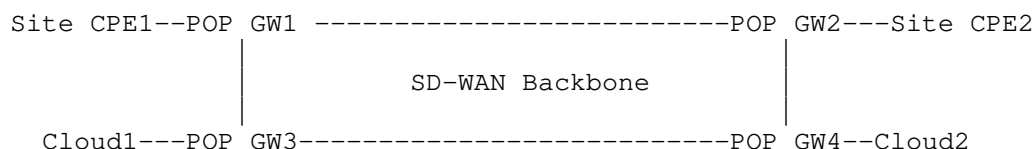


Figure 2. Typical topology of SD-WAN

Currently, the traffic from the CPE to POP GW is forwarded through the shortest forwarding path over the Internet, or an MPLS tunnel.

In addition, traffic from POP GW to another POP GW can be forwarded along with the MPLS tunnel that is a leased line, or over the internet, depending on the forwarding policies.

When the traffic is forwarded over the internet, it can be forwarded over a VXLAN tunnel. However, when using VXLAN, only the overlay connection is provided to enterprises/tenants, while the underlay forwarding path can not be specified and programmed. Therefore the SLA requirements can not be guaranteed when the traffic is forwarded overlay.

4. IPv6-based Cloud-Oriented Networking

This document describes a networking architecture called IPv6-based Cloud-Oriented Networking (CON). IPv6-based CON is an IPv6-based networking which provide quick, flexible connection to support dynamic site to cloud, and inter-cloud connections. Also, it supports end-to-end underlay forwarding path programming, so that services like strict path TE and SFC can be supported better.

The following section describes the requirements in IPv6-based CON, and the related solutions that meet the requirements.

4.1. Requirements

This section describes the overall requirements which need to be fulfilled by IPv6-based Cloud-Oriented Networking.

4.1.1. Quick Connection

Enterprise sites can locate at any location around the world, they need to connect to the clouds or other sites in any time, from any where. Also, enterprises may have some Virtual Private Clouds (VPC) in different clouds, they need to connect to each other in real time as well. The servers may locate in different cloud data centers or enterprise sites, which provide services for employees or other users. Therefore quick connection is required in IPv6-based CON.

4.1.2. Hybrid Network Connection

The enterprise VPN traffic can be forwarded around the world, which may travel heterogeneous networks, such as IPv4, MPLS and IPv6.

Typically, when a SD-WAN network connects multiple sites and clouds, it may cover hybrid networks. For example, the sub-path from the CPE to POP WG could be an IPv4 sub-path without any resource guarantee. The sub-path between POP GWs could be an MPLS LSP with resource reservation.

Therefore, connection over hybrid networks MUST be supported in IPv6-based CON.

4.1.3. Path Programming

When the enterprise VPN traffic is forwarded among sites or clouds, it may be forwarded along different paths. Each path has different performance such as different bandwidth, delay, etc. For instance, path A is the shortest path from site 1 to cloud 1, which has the lowest delay, while the path B can provide more bandwidth than path A. Therefore, the delay-sensitive traffic like PC gaming traffic SHOULD be forwarded along with path A, and the traffic that is delay-insensitive but requiring large bandwidth SHOULD be forwarded along with path B.

In order to meet the different SLA requirements, IPv6-based CON MUST support path programming.

4.1.4. Resource Assurance

In RSVP-TE MPLS, resources like bandwidth can be reserved for an LSP. When the traffic is forwarded along the LSP, the bandwidth can be guaranteed, which makes sure that the traffic will not be affected by other traffic. In order to provide SLA guaranteed services, IPv6-based CON MUST support Resource Assurance.

Network slicing is an approach to provide separate and independent end-to-end logical network over the physical network infrastructure. Each Network Slicing has its own resources, which can meet the specific SLA requirements. In order to provide SLA guaranteed services, IPv6-based CON MUST support network slicing.

4.1.5. Deterministic Delay

Delay-sensitive traffic has the strict requirements of network delay. Especially, when the servers moved to clouds instead of locating locally within the enterprise site, the long physical distance of packet forwarding path will introduce larger delay. In the traditional network, the shortest forwarding path is calculated based on the metric, and the metric is usually associated to the physical hops instead of latency. However, minimum delay forwarding is required for delay-sensitive traffic, like real-time video broadcast and video meeting.

Therefore, IPv6-based CON MUST have the capability to support path computing based on delay. Also, it MUST be able to provide deterministic delay forwarding.

4.1.6. Service Function Chaining

Service Function Chaining [RFC7665] is a mechanism to provide different value-added services (VAS) for packets.

A service function chain defines an ordered set of abstract service functions and ordering constraints that must be applied to packets and/or frames and/or flows selected as a result of classification [RFC7665].

An example of an abstract service function is "a firewall". Typically, different tenant's traffic in cloud data center will traverse different services function chain containing Firewall, DPI or other VAS.

Therefore, IPv6-based CON MUST have the capability to support SFC.

4.1.7. Performance Measurement

Many OAM mechanisms are used to support network operation. Performance Measurement (PM) is one of the most important part of OAM. With PM, the real-time QoS of the forwarding path, like delay, packet loss ratio and throughput, can be measured.

PM can be implemented in one of three ways: active, passive, or hybrid [RFC7799], differing in whether OAM packets need to be proactively sent.

On-path telemetry [I-D.song-opsawg-ifit-framework] is an hybrid mode OAM/PM mechanism, which provides better accuracy than active PM. Therefore, on-path Performance Measurement MUST be supported in IPv6-based CON.

4.1.8. Reliability

In Cloud-Network Interconnection scenarios, the enterprise traffic is forwarded over the WAN paths. The traffic can be sensitive to delay or packet losing, so high reliability is required in these scenarios. Therefore, protection of node and links MUST be supported in IPv6-based CON. Furthermore, redundancy transmission SHOULD be supported.

4.1.9. Security

As mentioned above, the enterprise traffic is forwarded over the WAN paths in IPv6-based CON. The security of the traffic MUST be ensured.

Also, in SD-WAN scenarios, customers are most concerned about security.

Therefore, IPv6-based CON MUST support secure connection, and MUST provide security assurance for the traffic in transmission.

4.1.10. Forwarding Efficiency

Tenants/Customers rent the physical or logical WAN links/paths from network operators for building they cloud-network interconnection enterprise network, so the forwarding efficiency is important for the WAN path tenant.

Path Maximum Transmission Unit indicates the maximum size of a packet that it can be forwarded along a path. Setting an appropriate PMTU for packets can avoid fragmentation or dropping, so that the forwarding efficiency can be raised.

Also, the overhead of packets MUST be added very carefully since it affects the forwarding efficiency directly. Especially, when many SIDs are inserted in an SRv6 packet, the overhead of the SID list is too large. [I-D.srcompdt-spring-compression-requirement] describes the requirements of SRv6 compression.

Therefore, the IPv6-based CON MUST support PMTU probing and configuration. In addition, it MUST support SRv6 compression.

4.1.11. Application-Aware Networking

Network operators are typically unaware of which applications are traversing their networks, which is because the network layer is decoupled from application layer. Adding application knowledge to the network layer enables finer granularity requirements of applications to be specified to the network operator. As IPv6 is being widely deployed, the programmability provided by IPv6 encapsulations can be augmented by conveying application information.

In IPv6-based CON, many types of applications' traffic is exchanged between sites and clouds. They have various requirements of QoS, and should be treated differently. In order to provide finer granularity traffic engineering to reduce the cost of WAN services, application-aware networking SHOULD be supported in IPv6-based CON.

4.2. Solutions

This section describes the candidate solutions that meet the requirements in IPv6-based CON.

4.2.1. VPN

VPN is a basic and essential services for cloud-networks interconnections.

SRv6 supports VPN by encoding the VPN information into the VPN SID [I-D.ietf-spring-srv6-network-programming].

Based on IPv6, SRv6 VPN can be established across multiple domains. It avoids configuring VPN services at each boundary nodes at each domain like the way in IPv4/MPLS networks (Option A). Deploying VPN based on SRv6 can shorten the duration significantly.

Also, L2VPN and L3VPN can be supported uniformly based on EVPN control plane [I-D.ietf-bess-srv6-services]. Therefore, combining the SRv6 data plane and EVPN control plane, the VPN can be deployed in an easy and flexible way in IPv6-based CON.

4.2.2. Path Programming

Based on SRv6, the traffic forwarding path can be programmed at the ingress of the SRv6 domain, so that the traffic from sites to clouds or inter-cloud can be forwarded through the specific underlay path.

For instance, in SD-WAN scenarios, the POP GW can choose a specific underlay forwarding path in WAN by choosing a binding SID [I-D.dukes-spring-sr-for-sdwan]. If the CPE supports SRv6, a controller can convey the programmed path information to the CPE via BGP SRv6 policy [I-D.ietf-idr-segment-routing-te-policy] or PCEP SRv6 policy [I-D.ietf-pce-segment-routing-policy-cp].

If the WAN connection travels multiple domains, the WAN path can be connected by multiple tunnels, such as VXLAN, GRE tunnel. [I-D.dunbar-sr-sdwan-over-hybrid-networks] describes how to associated the tunnels.

In order to shorten the delay, a CPE or PE can choose the nearest server in a specific cloud, and forward the packets through programmed paths.

4.2.3. Service Function Chaining

SFC is required in IPv6-based CON since different tenants subscribe different value-added services.

[I-D.ietf-spring-sr-service-programming] defines the mechanism to support SFC in SRv6. Each service function (SF) can be represented as an SRv6 SID if it supports SRv6. If the SF is SRv6-unaware device, then proxy SID is used. By programming service SIDs into the SRH, the SFC can be supported in SRv6.

Thanks to IPv6 reachability, SRv6 supports to program the end-to-end forwarding path from the carrier network to the inside the cloud data center, even to multiple clouds.

If NSH-based SFC has been deployed, a transition solution should be considered, and [I-D.ietf-spring-nsh-sr] describes a NSH and SR integration SFC solution.

4.2.4. IPv6 based Network Slicing

The tenant traffic MUST be isolated in WAN to avoid affecting by other internet traffic.

A framework, Enhanced VPN (VPN+), to form an enhanced connectivity services between customer sites is defined as per

[I-D.ietf-teas-enhanced-vpn]. Typically, VPN+ will be used to form the underpinning of network slicing. It also defines Virtual Transport Network (VTN). A VTN is a virtual underlay network that connects customer edge points with the capability of providing the isolation and performance characteristics required by an enhanced VPN customer. A VTN usually has a customized topology and a set of dedicated or shared network resources [I-D.ietf-teas-enhanced-vpn].

A VTN-ID option in IPv6 HBH header is defined in [I-D.dong-6man-enhanced-vpn-vtn-id] to identify the Virtual Transport Network (VTN) the packet belongs to. A VTN can be used as the underlay for one or a group of VPNs to provide enhanced VPN (VPN+) services.

By using VTN-ID, an end-to-end IPv6 network slicing is identified in transport network. Tenant traffic in WAN can be forwarded in the VTN with guaranteed resource.

4.2.5. IPv6-based On-path Measurement

The extension of supporting Alternate Marking Method [RFC8321] in IPv6 is defined in [I-D.ietf-6man-ipv6-alt-mark]. It describes how the Alternate Marking Method to be used as the hybrid performance measurement tool in an IPv6 domain by defining a new Extension Header Option.

Alternate Marking Method is a hybrid on-path performance measurement method, and the metadata of each node can be collected by the collector to compute the performance of the path.

IOAM is another on-path measurement method.

[I-D.ietf-ippm-ioam-ipv6-options] defines a new IPv6 option, called IOAM option to support carrying IOAM metadata in the IPv6 data packet. However, carrying all the metadata in the data packet will bring challenges for hardware processing. For instance, long-length metadata may cause recircle in processing. Therefore, [I-D.ietf-ippm-ioam-direct-export] defines a direct export option in IOAM, which enables the nodes to export the metadata to collector directly. Furthermore, [I-D.song-opsawg-ifu-framework] outlines a high-level framework to provide an operational environment that utilizes existing and emerging on-path telemetry techniques to enable the collection and correlation of performance information from the network.

4.2.6. Reliability

4.2.6.1. Local Protection

Local protection mechanisms like Fast Reroute (FRR) provide 50 ms protection on nodes for traffic.

Regarding link failures, TI-LFA

[I-D.ietf-rtgwg-segment-routing-ti-lfa] provides a fast reroute mechanism by sending the traffic to an expected post-convergence paths from the point of local repair.

Regarding the middle endpoint node failures,

[I-D.hu-spring-segment-routing-proxy-forwarding] defines a mechanism for fast reroute protection against the failure of a SR-TE path. It can provide fast reroute protection for an adjacency segment, a node segment and a binding segment of the path. Also,

[I-D.chen-rtgwg-srv6-midpoint-protection] defines midpoint protection, which enables the direct neighbor of the failed endpoint to perform the function of the endpoint, replace the IPv6 destination address to the next endpoint, and choose the next hop based on the new destination address.

Regarding the egress node failures,

[I-D.ietf-rtgwg-srv6-egress-protection] defines a local protection solution using the mirror SID.

4.2.6.2. End-to-End Protection

End-to-End Protection is also required in IPv6-based CON. Normally, host-standby nodes are deployed for supporting traffic switching from the failed node to the alternative node. In order to detect the failure, End-to-end BFD is required. Once the BFD session is failed, the traffic can be steered into a disjoint forwarding path, and the traffic will be forwarded to the host-standby node.

4.2.6.3. Redundancy Protection

In order to avoid losing packets,

[I-D.geng-spring-sr-redundancy-protection] defines a redundancy transmission solution.

The document defines two types of segment including Redundancy Segment and Merging Segment to empower the Segment Routing with the capability of redundancy protection.

4.2.7. Security

As per [I-D.li-spring-srv6-security-consideration], SRv6 inherits potential security vulnerabilities from Source Routing and IPv6, but it does not introduce new critical security threats.

Regarding a limited domain, SPRING architecture [RFC8402] defines clear trusted domain boundaries so that source-routing information is only available within the trusted domain and never exposed to the outside of the domain. It is expected that, by default, explicit routing is only used within the boundaries of the administered domain. Therefore, the data plane does not expose any source-routing information when a packet leaves the trusted domain. The traffic is filtered at the domain boundaries [RFC8402].

However, it has been noted that the AH and ESP are not directly applicable in order to reduce the vulnerabilities of SRv6 due to the presence of mutable fields in the SRH [I-D.li-spring-srv6-security-consideration]. In order to resolve this problem, [RFC8754] defines a mechanism to carry HMAC TLV in the SRH to verify the integrity of packets including the SRH fields.

Regarding end-to-end security protection across multiple domains, an end-to-end IPsec tunnel is suggested to be deployed.

In typical SD-WAN scenarios, the IPsec tunnel should be used between the CPE and POP GW.

4.2.8. IPv6 Forwarding Efficiency

4.2.8.1. PMTU

The host may discover the PMTU by Path MTU Discovery (PMTUD) [RFC8201] or other means. But the ingress node still needs to examine the packet size to drop too large packets to avoid malicious packets or error packets attack. Also, the packet size may exceed the PMTU because of the new encapsulation of SR-MPLS or SRv6 at the ingress. In order to check whether the packet size exceeds the PMTU or not, the ingress node need to know the Path MTU associated to the forwarding path.

However, the path maximum transmission unit (MTU) information for SR path is not available since the SR does not require signaling. [I-D.ietf-idr-bgp-ls-link-mtu] proposes a BGP-LS extensions to collect the link MTU of the links in the networks. [I-D.ietf-idr-sr-policy-path-mtu] and [I-D.li-pce-pcep-pmtu] defines extensions to distribute path MTU information within BGP and PCEP SR

policies, respectively. In this way, the controller can compute the appropriate PMTU for an SR path.

4.2.8.2. SRv6 Compression

The overhead of SRv6 encapsulation brings challenges for hardware processing and transmission.

[I-D.srcompdt-spring-compression-requirement] describes the requirements of SRv6 compression.

G-SRv6 is proposed in [I-D.cl-spring-generalized-srv6-np], which supports to encode multiple types of SIDs in SRH. This SRH is called Generalized SRH [I-D.lc-6man-generalized-srh] while the SID is called Generalized SID.

G-SRv6 supports to encode the compressed SIDs in the SRH to reduce the size of SRv6 SID list in SRH

[I-D.cl-spring-generalized-srv6-for-cmpr]. A COC (Continuation of Compression) flavor is defined to indicate the continuation of SRv6 compressed SIDs in the SID list.

4.2.9. Application-aware IPv6 Networking

Application-aware Networking (APN) is proposed by [I-D.li-apn-framework], where application characteristic information such as application identification and its network performance requirements is carried in the packet encapsulation in order to facilitate service provisioning, perform application-level traffic steering and network resource adjustment.

Application-aware IPv6 Networking (APN6) framework makes use of IPv6 encapsulation in order to convey the application-aware information along with the data packet to the network so to facilitate the service deployment and SLA guarantee.

[I-D.li-6man-app-aware-ipv6-network] defines the encodings of the application characteristic information, for the APN6 framework, that can be exchanged between an application and the network infrastructure through IPv6 extension headers.

5. IANA Considerations

TBD

6. Security Considerations

TBD

7. Contributors

TBD

8. Acknowledgements

9. References

9.1. 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>>.
- [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] Filshill, 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>>.
- [RFC8402] Filshill, 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>>.

9.2. Informative References

- [I-D.ietf-rtgwg-net2cloud-problem-statement] Dunbar, L., Malis, A., 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.

- [I-D.ietf-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., Voyer, D.,
Matsushima, S., and Z. Li, "SRv6 Network Programming",
draft-ietf-spring-srv6-network-programming-28 (work in
progress), December 2020.
- [I-D.ietf-bess-srv6-services]
Dawra, G., Filsfils, C., Talaulikar, K., Raszuk, R.,
Decraene, B., Zhuang, S., and J. Rabadan, "SRv6 BGP based
Overlay services", draft-ietf-bess-srv6-services-05 (work
in progress), November 2020.
- [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>>.
- [RFC7799] Morton, A., "Active and Passive Metrics and Methods (with
Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799,
May 2016, <<https://www.rfc-editor.org/info/rfc7799>>.
- [I-D.dukes-spring-sr-for-sdwan]
Dukes, D., Filsfils, C., Dawra, G., Xu, X., Voyer, D.,
Camarillo, P., Clad, F., and S. Salsano, "SR For SDWAN:
VPN with Underlay SLA", draft-dukes-spring-sr-for-sdwan-02
(work in progress), June 2019.
- [I-D.dunbar-sr-sdwan-over-hybrid-networks]
Dunbar, L. and M. Toy, "Segment routing for SDWAN paths
over hybrid networks", draft-dunbar-sr-sdwan-over-hybrid-
networks-06 (work in progress), November 2019.
- [I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P.,
Rosen, E., Jain, D., and S. Lin, "Advertising Segment
Routing Policies in BGP", draft-ietf-idr-segment-routing-
te-policy-11 (work in progress), November 2020.
- [I-D.ietf-pce-segment-routing-policy-cp]
Koldychev, M., Sivabalan, S., Barth, C., Peng, S., and H.
Bidgoli, "PCEP extension to support Segment Routing Policy
Candidate Paths", draft-ietf-pce-segment-routing-policy-
cp-02 (work in progress), January 2021.

- [I-D.ietf-spring-sr-service-programming]
Clad, F., Xu, X., Filsfils, C., daniel.bernier@bell.ca,
d., Li, C., Decraene, B., Ma, S., Yadlapalli, C.,
Henderickx, W., and S. Salsano, "Service Programming with
Segment Routing", draft-ietf-spring-sr-service-
programming-03 (work in progress), September 2020.
- [I-D.ietf-spring-nsh-sr]
Guichard, J. and J. Tantsura, "Integration of Network
Service Header (NSH) and Segment Routing for Service
Function Chaining (SFC)", draft-ietf-spring-nsh-sr-04
(work in progress), December 2020.
- [I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A
Framework for Enhanced Virtual Private Networks (VPN+)
Service", draft-ietf-teas-enhanced-vpn-06 (work in
progress), July 2020.
- [I-D.dong-6man-enhanced-vpn-vtn-id]
Dong, J., Li, Z., Xie, C., and C. Ma, "Carrying Virtual
Transport Network Identifier in IPv6 Extension Header",
draft-dong-6man-enhanced-vpn-vtn-id-02 (work in progress),
November 2020.
- [RFC8321] Fioccola, G., Ed., Capello, A., Cociglio, M., Castaldelli,
L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi,
"Alternate-Marking Method for Passive and Hybrid
Performance Monitoring", RFC 8321, DOI 10.17487/RFC8321,
January 2018, <<https://www.rfc-editor.org/info/rfc8321>>.
- [I-D.ietf-6man-ipv6-alt-mark]
Fioccola, G., Zhou, T., Cociglio, M., Qin, F., and R.
Pang, "IPv6 Application of the Alternate Marking Method",
draft-ietf-6man-ipv6-alt-mark-02 (work in progress),
October 2020.
- [I-D.ietf-ippm-ioam-ipv6-options]
Bhandari, S., Brockners, F., Pignataro, C., Gredler, H.,
Leddy, J., Youell, S., Mizrahi, T., Kfir, A., Gafni, B.,
Lapukhov, P., Spiegel, M., Krishnan, S., Asati, R., and M.
Smith, "In-situ OAM IPv6 Options", draft-ietf-ippm-ioam-
ipv6-options-04 (work in progress), November 2020.

- [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", draft-ietf-ippm-ioam-direct-
export-02 (work in progress), November 2020.
- [I-D.song-opsawg-ifit-framework]
Song, H., Qin, F., Chen, H., Jin, J., and J. Shin, "In-
situ Flow Information Telemetry", draft-song-opsawg-ifit-
framework-13 (work in progress), October 2020.
- [I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B.,
and D. Voyer, "Topology Independent Fast Reroute using
Segment Routing", draft-ietf-rtgwg-segment-routing-ti-
lfa-05 (work in progress), November 2020.
- [I-D.hu-spring-segment-routing-proxy-forwarding]
Hu, Z., Chen, H., Yao, J., Bowers, C., and Y. Zhu, "SR-TE
Path Midpoint Protection", draft-hu-spring-segment-
routing-proxy-forwarding-12 (work in progress), October
2020.
- [I-D.chen-rtgwg-srv6-midpoint-protection]
Chen, H., Hu, Z., Chen, H., and X. Geng, "SRv6 Midpoint
Protection", draft-chen-rtgwg-srv6-midpoint-protection-03
(work in progress), October 2020.
- [I-D.ietf-rtgwg-srv6-egress-protection]
Hu, Z., Chen, H., Chen, H., Wu, P., Toy, M., Cao, C., He,
T., Liu, L., and X. Liu, "SRv6 Path Egress Protection",
draft-ietf-rtgwg-srv6-egress-protection-02 (work in
progress), November 2020.
- [I-D.geng-spring-sr-redundancy-protection]
Geng, X., Chen, M., and F. Yang, "Segment Routing for
Redundancy Protection", draft-geng-spring-sr-redundancy-
protection-00 (work in progress), November 2020.
- [I-D.li-spring-srv6-security-consideration]
Li, C., Li, Z., Xie, C., Tian, H., and J. Mao, "Security
Considerations for SRv6 Networks", draft-li-spring-srv6-
security-consideration-05 (work in progress), October
2020.

- [RFC8201] McCann, J., Deering, S., Mogul, J., and R. Hinden, Ed., "Path MTU Discovery for IP version 6", STD 87, RFC 8201, DOI 10.17487/RFC8201, July 2017, <<https://www.rfc-editor.org/info/rfc8201>>.
- [I-D.ietf-idr-bgp-ls-link-mtu] Zhu, Y., Hu, Z., Peng, S., and R. Muehler, "Signaling Maximum Transmission Unit (MTU) using BGP-LS", draft-ietf-idr-bgp-ls-link-mtu-00 (work in progress), November 2020.
- [I-D.ietf-idr-sr-policy-path-mtu] Li, C., Zhu, Y., Sawaf, A., and Z. Li, "Segment Routing Path MTU in BGP", draft-ietf-idr-sr-policy-path-mtu-02 (work in progress), November 2020.
- [I-D.li-pce-pcep-pmtu] Peng, S., Li, C., Han, L., and L. Ndifor, "Support for Path MTU (PMTU) in the Path Computation Element (PCE) communication Protocol (PCEP).", draft-li-pce-pcep-pmtu-03 (work in progress), October 2020.
- [I-D.srcompdt-spring-compression-requirement] Cheng, W., "Compressed SRv6 SID List Requirements", draft-srcompdt-spring-compression-requirement-03 (work in progress), January 2021.
- [I-D.cl-spring-generalized-srv6-np] Cheng, W., Li, Z., Li, C., Xie, C., Li, C., Tian, H., and F. Zhao, "Generalized SRv6 Network Programming", draft-cl-spring-generalized-srv6-np-02 (work in progress), September 2020.
- [I-D.lc-6man-generalized-srh] Li, Z., Li, C., Cheng, W., Xie, C., Cong, L., Tian, H., and F. Zhao, "Generalized Segment Routing Header", draft-lc-6man-generalized-srh-01 (work in progress), August 2020.
- [I-D.cl-spring-generalized-srv6-for-cmpr] Cheng, W., Li, Z., Li, C., Clad, F., Aihua, L., Xie, C., Liu, Y., and S. Zadok, "Generalized SRv6 Network Programming for SRv6 Compression", draft-cl-spring-generalized-srv6-for-cmpr-02 (work in progress), November 2020.

[I-D.li-apn-framework]

Li, Z., Peng, S., Voyer, D., Li, C., Geng, L., Cao, C.,
Ebisawa, K., Previdi, S., and J. Guichard, "Application-
aware Networking (APN) Framework", draft-li-apn-
framework-01 (work in progress), September 2020.

[I-D.li-6man-app-aware-ipv6-network]

Li, Z., Peng, S., Li, C., Xie, C., Voyer, D., Li, X., Liu,
P., Liu, C., and K. Ebisawa, "Application-aware IPv6
Networking (APN6) Encapsulation", draft-li-6man-app-aware-
ipv6-network-02 (work in progress), July 2020.

Authors' Addresses

Cheng Li (editor)
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: c.l@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com

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

Email: hongjie.yang@huawei.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 23, 2021

M. McBride
J. Guichard
Y. Qu
Futurewei
T. Hardjono
MIT
CJ. Bernardos
UC3M
February 19, 2021

Data Discovery Use Cases
draft-mcbride-data-discovery-use-cases-00

Abstract

There needs to be a solution for locating and capturing data in a standardized way. Data may be cached, copied and/or stored at multiple locations in the network on route to its final destination. With an increasingly high volume of devices connecting to the Internet, support for network caching and replication is critical for continuous data availability. There are data repositories throughout a modern network and there needs to be a standardized way to locating the repositories and discovering the desired data within.

There are several use cases which illustrate a need for a data discovery solution. An application might need to query the network to discover resources (program, service, resource) that can help the local application perform a particular task. Additionally, there could be volumes of data which needs to be searched and discovered in order to provide a result to be acted upon by the application. These are a couple of the use cases being addressed 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 August 23, 2021.

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. Terminology	3
4. Problem Statement	3
4.1. Types of Data	4
5. Use Cases	4
5.1. Application-Aware Service Function Chaining	4
5.2. Available CPU and Memory Resources	5
5.3. Data Dependency	5
5.4. Distributed Ledgers	5
5.5. Edge Computing	6
6. IANA Considerations	6
7. Security Considerations	6
8. Acknowledgements	6
9. Normative References	6
Authors' Addresses	7

1. Introduction

An application might need to query the network to discover resources that can help the local application perform a particular task. There could be volumes of data which needs to be searched and discovered in order to provide a result to be acted upon.

Data discovery might involve an application requesting data. It might involve a device looking to store data or to request the processing from a data store and then gather the result. Or it could be execution of a set of instructions at an appropriate device in the network. Another possible area is service chaining where an

application needs to run its data through a firewall but the selected firewall must have a particular rule set applicable to this particular application. Perhaps the service function has to be located within a particular environment (security level). Or a particular device must be found that is capable of executing upon a set of instructions provided in the data packet. This document focuses on various data discovery use cases.

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].

3. Terminology

- o SFC: Service Function Chaining
- o APN: Application-Aware Networking
- o DLT: Distributed Ledger Technologies

4. Problem Statement

As discussed in [I-D.mcbride-data-discovery-problem-statement], there are many proprietary and standardized ways of discovering networking devices and hosts. There are many solutions for discovering data within a database. There are proprietary, non-standardized, ways of discovering the data that may be stored throughout an environment of networking devices. We can discover information about the devices but can't locate and capture stored data (resource, program, service, etc) in a standard way. With more networking devices storing collected data there needs to be a standard way of discovering the specific data needed amongst a potentially huge lake of databases.

This data discovery problem is particularly true for use cases where it will be important to have the capability to express a data request within the data packets and have the network route the traffic accordingly. This might be an application requesting data. It might be a device looking to store data or to request the processing, and result, from a data store. It could be execution of a set of instructions at an appropriate device in the network. An application may need to run its data through a firewall but the selected firewall must have a particular rule set applicable to this particular application. Perhaps a service function needs to be located within a particular environment (security level). Or a particular device must be found that is capable of executing upon a set of instructions

provided in the data packet. This document focuses on data discovery use cases.

4.1. Types of Data

Discoverable data can be a resource, program, service etc. And an infinite amount, and types, of data can be discoverable including statistics, measurements, temperature, location, metadata, health, transactions and so on.

Program: applets, graphics, games, spreadsheets, database systems, browsers, etc

Service: firewalls, load balancers, spam filters, header manipulators, etc

Resource: CPU, memory, etc

5. Use Cases

Here are some use cases to illustrate the need for data discovery:

5.1. Application-Aware Service Function Chaining

Application Aware Networking (APN), as described in [I-D.li-apn-problem-statement-usecases], allows applications to specify finer granularity requirements to the network operator by providing application knowledge to the network layer. This granularity includes the ability to convey the characteristics of an application's traffic flow and program the network infrastructure accordingly to provide service assurance.

An application might need to query the network to discover resources that can help the local application perform a particular task. Additionally, there could be volumes of data which needs to be searched and discovered in order to provide a result to be acted upon by the application.

End-to-end service delivery often needs to go through various service functions, including traditional network service functions such as firewalls, DPIs as well as new application-specific functions, both physical and virtual. APN provides assigning a given traffic flow to a specific service function chain (SFC) but also specifically allows the subsequent steering according to the application information carried in the APN packets.

When an application needs to run its data through a firewall, but the selected firewall must have a particular rule set applicable to this

particular application, then the application can leverage data discovery functionality. The service function may be required to be located within a particular environment such as a with a certain security level. Data discovery is needed to find that particular rule set (amongst the various firewalls) and then steer the packet accordingly. Or a particular device, along the SFC, may need to be found that is capable of executing upon a set of instructions provided in the data packet. The data capabilities of devices needs to be discoverable in order to steer the application packets towards them along a SFC.

5.2. Available CPU and Memory Resources

An application, or service, may need to discover the available server memory and compute resources from the network. A certain amount of CPU resources may be required to support a particular application workload. And the application may need to know the maximum CPU utilization threshold available on a compute device. Gathering info on available clock speeds and amount of cores can help determine how quickly servers load and interact with a set of applications. The network can provide the discoverability of the necessary data (cpu, memory) in order for applications to properly execute. A network planning app can also utilize this information to help predict future resource demands in order to meet applications performance requirements.

5.3. Data Dependency

There may be scenarios where it's critical to find X type of data that can help a local application, or service, successfully perform a particular task. Perhaps an industrial application needs real time measurement data, such as temperature, in order to execute a process. This required data may be cached, copied and/or stored at multiple locations in the network on route to its final destination. With an increasing percentage of devices connecting to the Internet being mobile, support for in-the-network caching and replication is critical for continuous data availability, not to mention efficient network and battery usage for endpoint devices. In order for some applications to properly execute, we need to find a way for the network to provide support for data discovery.

5.4. Distributed Ledgers

DLT Gateways, as discussed in [I-D.sardon-blockchain-gateways-usecases], will be given a permissioned view of assets/transactions, that they are requested to transfer, within their attached DLT domain. GW's may also need to discover assets/transactions, not explicitly provided, within the DLT

domain. It may become necessary for the GW (or other network element.. if permitted) to discover the data (asset, resource, service...) in order to transfer the required asset. Discovery of the data parts is also needed to validate the transfer after the asset movement. The ledger in the DLT will not hold all the relevant information pertaining to a previous asset transfer. So there needs to be ways to search/discover these. The data parts, to be discovered, include:

Relevant DLT transaction public-keys of the involved entities (i.e. public-keys (addresses) used on both DLTs).

Relevant entity public-keys and X.509 certs (Originator, owner of gateway G1, owner of gateway G2, Beneficiary). This is similar to the X.509 certs and cert-profiles used in the SWIFT banking network.

Relevant asset-related JSON documents (e.g. asset profiles).

5.5. Edge Computing

As described in [I-D.mcbride-edge-data-discovery-overview], the required data may be distributed across thousands of edge computing devices. Edge computing is motivated by the sheer volume of data that is being created by endpoint devices (sensors, cameras, lights, vehicles, drones, wearables, etc.) at the very network edge. In dense IoT deployments (e.g., many video cameras are streaming high definition video), where multiple data flows collect or converge at edge nodes, data is likely to need transformation (transcoded, subsampled, compressed, analyzed, annotated, combined, aggregated, etc.) to fit over the next hop link, or even to fit in memory or storage. This data, distributed across the edge, will need to be discovered in order to perform any number of functions such as an IoT application needing elevator vibration data in order to execute a process.

6. IANA Considerations

7. Security Considerations

8. Acknowledgements

9. Normative References

[I-D.li-apn-problem-statement-usecases]

Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z.,
Ebisawa, K., Previdi, S., and J. Guichard, "Problem
Statement and Use Cases of Application-aware Networking
(APN)", draft-li-apn-problem-statement-usecases-01 (work
in progress), September 2020.

[I-D.mcbride-data-discovery-problem-statement]

McBride, M., Kutscher, D., Schooler, E., Bernardos, C.,
and D. Lopez, "Data Discovery Problem Statement", draft-
mcbride-data-discovery-problem-statement-00 (work in
progress), July 2020.

[I-D.mcbride-edge-data-discovery-overview]

McBride, M., Kutscher, D., Schooler, E., Bernardos, C.,
Lopez, D., and X. Foy, "Edge Data Discovery for COIN",
draft-mcbride-edge-data-discovery-overview-05 (work in
progress), November 2020.

[I-D.sardon-blockchain-gateways-usecases]

Sardon, A. and T. Hardjono, "Blockchain Gateways: Use-
Cases", draft-sardon-blockchain-gateways-usecases-00 (work
in progress), October 2020.

[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

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Jim Guichard
Futurewei

Email: james.n.guichard@futurewei.com

Yingzhen Qu
Futurewei

Email: yingzhen.qu@futurewei.com

Thomas Hardjono
MIT

Email: hardjono@mit.edu

Carlos J. Bernardos
Universidad Carlos III de Madrid
Av. Universidad, 30
Leganes, Madrid 28911
Spain

Phone: +34 91624 6236
Email: cjbc@it.uc3m.es
URI: <http://www.it.uc3m.es/cjbc/>

RTGWG Working Group
Internet-Draft
Intended status: Standards Track
Expires: 30 September 2021

G. Mirsky
X. Min
ZTE Corp.
G. Mishra
Verizon Inc.
29 March 2021

Integrated Operation, Administration, and Maintenance
draft-mmm-rtgwg-integrated-oam-01

Abstract

This document describes the Integrated Operation, Administration, and Maintenance (IntOAM) protocol. IntOAM is based on the lightweight capabilities of Bidirectional Forwarding Detection defined in RFC 5880 Bidirectional Forwarding Detection, and the RFC 6374 Packet Loss and Delay Measurement for MPLS Networks to measure performance metrics like packet loss and packet delay. Also, a method to perform lightweight on-demand authentication is defined in this specification.

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 30 September 2021.

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. Conventions used in this document	3
2.1. Acronyms	3
2.2. Requirements Language	3
3. Integrated OAM Control Message	3
4. Theory of Operation	6
4.1. Use of Discriminators	6
4.2. Modes of IntoOAM	7
4.3. Echo Function	7
5. Using TLVs in the IntoOAM	7
5.1. Integrated OAM Capability Negotiation	7
5.1.1. Timer Negotiation for Performance Monitoring	9
5.2. Padding TLV	10
5.3. Diagnostic TLV	11
5.4. Performance Measurement with IntoOAM Control Message	12
5.5. Lightweight Authentication	13
5.5.1. Lightweight Authentication Mode Negotiation	14
5.5.2. Using Lightweight Authentication	15
6. IANA Considerations	16
6.1. IntoOAM Message Types	16
6.2. Lightweight Authentication Modes	17
6.3. Return Codes	18
7. Security Considerations	18
8. Acknowledgements	19
9. References	19
9.1. Normative References	19
9.2. Informative References	19
Authors' Addresses	20

1. Introduction

[RFC5880] has provided the base specification of Bidirectional Forwarding Detection (BFD) as the light-weight mechanism to monitor a path continuity between two systems and detect a failure in the data-plane. Since its introduction, BFD has been broadly deployed. There were several attempts to introduce new capabilities in the protocol, some more successful than others. One of the obstacles to extending BFD capabilities may be seen in the compact format of the BFD control message. This document introduces the Integrated Operation, Administration, and Maintenance (IntoOAM) protocol based on BFD's lightweight capabilities. It uses informational elements defined in

[RFC6374] to measure various performance metrics, e.g., synthetic packet loss or packet delay. Combination of both Fault Management (FM) Performance Monitoring (PM) OAM functions in the IntOAM protocol is beneficial in some networks. For example, in a Deterministic Networking (DetNet) domain [RFC8655], it is easier to ensure that an IntOAM's test packet is fate-sharing with data packets rather than mapping several FM and PM OAM protocols to that DetNet data flow.

2. Conventions used in this document

2.1. Acronyms

BFD: Bidirectional Forwarding Detection

G-ACh Generic Associated Channel

IntOAM Integrated OAM

HMAC Hashed Message Authentication Code

MTU Maximum Transmission Unit

PMTUD Path MTU Discovery

PMTUM Path MTU Monitoring

p2p: Point-to-Point

TLV Type, Length, Value

OAM Operations, Administration, and Maintenance

FM Fault Management

PM Performance Monitoring

DetNet Deterministic Networking

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

3. Integrated OAM Control Message

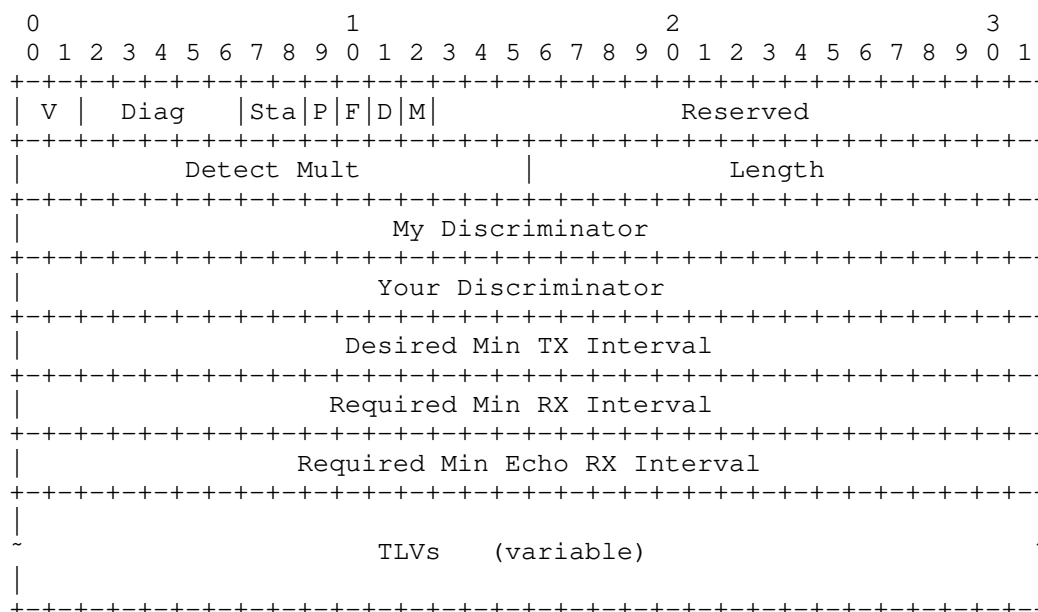


Figure 1: Integrated OAM Control Message Format

where fields are defined as the following:

- * Version (V) - two-bit field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the Version field.
- * Diagnostic (Diag) - five-bit field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the Diagnostic field.
- * Status (Sta) - two-bit field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the Status field.
- * Poll (P) - one-bit field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the Poll field.
- * Final (F) - one-bit field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the Final field.

- * Demand (D) - one-bit field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the Demand field.
- * Multipoint (M) - one-bit field. The definition of the field, its interpretation, and its use in the protocol operation are as defined in [RFC5880] for the Multipoint field.
- * Reserved - seventeen-bit field that can be defined in the future. It MUST be zeroed on transmission and ignored on receipt.
- * Detect Mult - two-octet field. The definition of the field, its interpretation, and its use in the protocol operation are as defined in [RFC5880] for the Detect Mult field.
- * Length - two-octet field equal to the length of the IntoAM Control message in octets.
- * My Discriminator - four-octet field. The definition of the field, its interpretation, use in the protocol operation, and assigned values are as defined in [RFC5880] for the My Discriminator field.
- * Your Discriminator - four-octet field. The definition of the field, its interpretation, and its use in the protocol operation are as defined in [RFC5880] for the Your Discriminator field.
- * Desired Min TX Interval - four-octet field. The definition of the field, its interpretation, and its use in the protocol operation are as defined in [RFC5880] for the Desired Min TX Interval field. Additional use cases for the Desired Min TX Interval field described in Section 5.1.1.
- * Required Min RX Interval - four-octet field. The definition of the field, its interpretation, and its use in the protocol operation are as defined in [RFC5880] for the Required Min RX Interval field. Additional use cases for the Required Min RX Interval field described in Section 5.1.1.
- * Required Min Echo RX Interval - four-octet field. [Ed.note: In BFD, as I understand, it serves several purposes - indicate support of Echo (zero value - non-support), and throttle rate the remote will send its Echo. But that only works if the Echo can be sent when the session is Up. There's now a proposal to send Echo regardless of the state of a session. Hence the question - is it still a good use of four bytes?]
- * TLVs - is a variable-length field that contains commands and/or data encoded as type-length-value (TLV) shown in Figure 2.

TLV is a variable-length field. Multiple TLVs MAY be placed in an IntoAM Control message. Additional TLVs may be enclosed within a given TLV, subject to the semantics of the (outer) TLV in question. If more than one TLV is to be included, the value of the Type field of the outmost outer TLV MUST be set to Multiple TLVs Used (TBA0), as assigned by IANA according to Section 6.1. Figure 2 displays the TLV format in an IntoAM protocol.

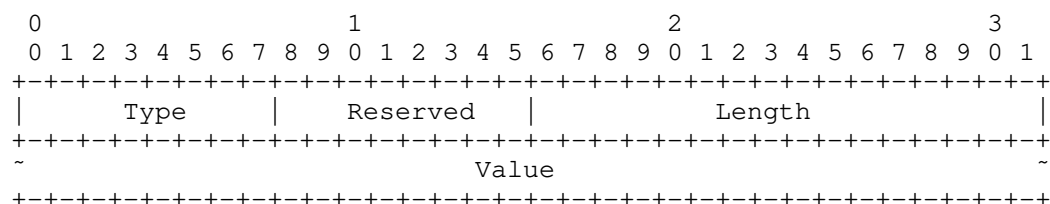


Figure 2: General Type-Length-Value Encoding

where fields are defined as the following:

- * Type - one-octet field that characterizes the interpretation of the Value field. Type values are allocated according to Section 6.1.
- * Reserved - one-octet field. The value of the Type field determines its interpretation and encoding.
- * Length - two-octet field equal to the length of the Value field in octets.
- * Value - a variable-length field. The value of the Type field determines its interpretation and encoding.

4. Theory of Operation

[Ed.note: Should the document reference Asynchronous and Demand modes in RFC 5880?]

4.1. Use of Discriminators

A discriminator is defined in the IntoAM as an unsigned 32-bit long integer that identifies a particular IntoAM session. An IntoAM system MAY locally assign a discriminator for the given IntoAM session. Also, a discriminator MAY be distributed by the control plane or management plane.

In a point-to-point (p2p) IntoAM session, the value of the Your Discriminator field is used to demultiplex IntoAM sessions. An IntoAM system has to learn the value of discriminator that the remote IntoAM system associates with the IntoAM session between these two systems. The IntoAM system MAY use a three-way handshake mechanism to learn of discriminator of the remote system. Besides, the control or management plane MAY be used to associate discriminator values with the specific IntoAM session. In other scenarios, e.g., point-to-multipoint (p2mp) IntoAM session, the Your Discriminator's value could be left undefined for some nodes. In that case, such a node uses the My Discriminator field's value in combination with information that identifies the sender of the IntoAM Control message and the path identifier.

4.2. Modes of IntoAM

IntoAM has two operational modes that provide for proactive defect detection in a network- Asynchronous and Demand. An IntoAM implementation MUST be capable of operating in either of them. In the Asynchronous mode, an IntoAM system periodically transmits IntoAM Control messages. When an IntoAM system is in the Demand mode, it does not periodically transmit IntoAM Control messages. An IntoAM system in the Demand mode MAY transmit a Control message as a part of the Poll sequence. A system MAY be set into the Demand mode at any time during the IntoAM session.

4.3. Echo Function

The Echo function in IntoAM can be used in networks when an operator has ensured that the sender's test packet will first reach the intended target before being returned to the sender. The target node is not required to support IntoAM as the IntoAM packet is expected to be looped back by the data plane without the need to inspect the test packet itself. The IntoAM Control message and IntoAM TLVs MAY be used as the test packet by the IntoAM Echo function.

5. Using TLVs in the IntoAM

5.1. Integrated OAM Capability Negotiation

An IntoAM system, also referred to as a node in this document, that supports IntoAM first MUST discover the extent to which other nodes in the given session support the Integrated OAM. The node MUST send an IntoAM Control message initiating the Poll Sequence as defined in [RFC5880]. If the remote system fails to respond with the IntoAM Control message and the Final flag set, then the initiator node MUST conclude that the peer does not support the use of the IntoAM Control messages.

The first IntoOAM Control message initiating the Poll Sequence SHOULD include the Capability TLV that lists capabilities that may be used at some time during the lifetime of the IntoOAM session. Until the node negotiated the use of PM capabilities of the IntoOAM, the node MUST NOT include any TLVs in the IntoOAM Control message, other than the Capability TLV. The format of the Capability TLV is presented in Figure 3.

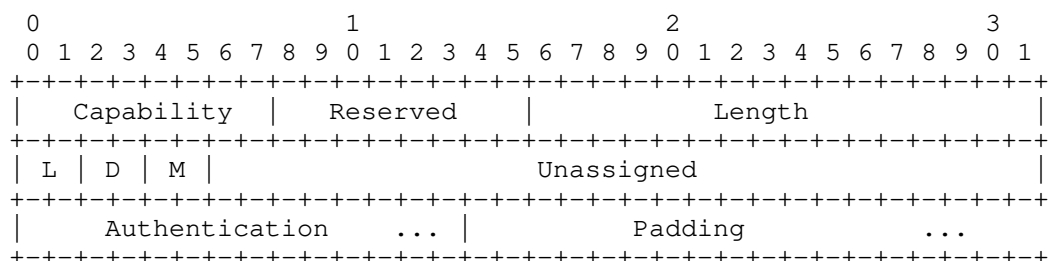


Figure 3: Capability TLV Format

where fields are defined as the following:

- * Capability - one-octet field. Its value (TBA2) allocated by IANA in Section 6
- * Reserved - one-octet field. It MUST be zeroed on the transmit and ignored on the receipt.
- * Length - two-octet field. The value equals length on the Capability TLV in octets. The value of the Length field MUST be a multiple of 4.
- * Loss - two-bit field. The least significant of the two bits is set if the node can measure packet loss using a periodically transmitted IntoOAM control message. The most significant of the two bits is set if the node is capable of measuring packet loss using the Poll Sequence with IntoOAM Control message.
- * Delay - two-bit field. The least significant of the two bits is set if the node can measure packet delay using a periodically transmitted IntoOAM control message. The most significant of the two bits is set if the node is capable of measuring packet delay using the Poll Sequence with IntoOAM Control message.
- * MTU - two-bits field. Set if the node is capable of using the IntoOAM Control message in Path MTU Discovery (PMTUD). or PMTU Monitoring (PMTUM). The least significant of the two bits is set

if the node can perform PMTUD/PMTUM using periodically transmitted IntoOAM control message. The most significant of the two bits is set if the node is capable of PMTUD/PMTUM using the Poll Sequence with IntoOAM Control message.

- * Unassigned - 26-bit field. It MUST be zeroed on transmission and ignored on receipt
- * (Lightweight) Authentication - variable-length field. An IntoOAM system uses the Authentication field for advertising its lightweight authentication capabilities. The format and the use of the Authentication field are defined in Section 5.5.1.
- * Padding - variable-length field. It MUST be zeroed on transmission and ignored on receipt. The Padding field aligns the length of the Capability TLV to a four-octet boundary.

The remote IntoOAM node that supports this specification MUST respond to the Capability TLV with the IntoOAM Control message, including the Capability TLV listing capabilities the responder supports. The responder MUST set the Final flag in the IntoOAM Control message.

5.1.1. Timer Negotiation for Performance Monitoring

IntoOAM allows for the negotiation of time intervals at which an IntoOAM system transmits and receives IntoOAM Control packets. That equally applies to packets used for performance monitoring, whether it measures packet delay or packet loss, using TLVs defined in Section 5.4. An IntoOAM system sets its timer values in an IntoOAM Control packet that includes the Capabilities TLV. The negotiation process is similar to the one described in [RFC5880]. A local IntoOAM system advertises its shortest interval for transmitting IntoOAM packets to measure the indicated metrics and the shortest interval that is it capable of receiving PM IntoOAM packets. Suppose a system does not support the given metric measurement, i.e., packet loss or packet delay. In that case, it MUST set the value of the Required Min RX Interval to zero when transmitting the IntoOAM Control message with the Capability TLV. If an IntoOAM system does not support one of the modes, periodic or on-demand, for the given performance metric, it MUST zero the appropriate bit in the field that describes the metric. The timer values apply to all PM modes that have their respective bits set in the Capacity TLV. If an operator wants to use a different time intervals for different performance metrics measurements, then separate Poll sequences with the Capabilities TLV included MAY be used. Thus IntoOAM allows negotiating different time intervals for packet loss and packet delay measurements.

5.2. Padding TLV

Padding TLV MAY be used to generate IntoAM Control messages of the desired length.

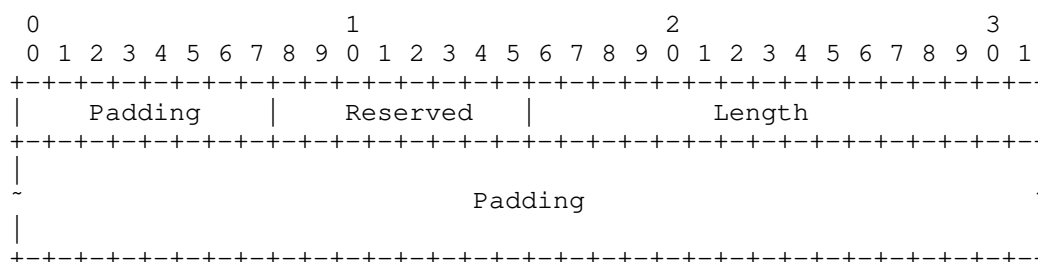


Figure 4: Padding TLV Format

where fields are defined as the following:

- * Padding - one-octet field. Its value (TBA1) allocated by IANA in Section 6
- * Reserved - one-octet field. MUST be zeroed on the transmit and ignored on the receipt.
- * Length - two-octet field equals length on the Padding TLV in octets. The value of the Length field MUST be a multiple of 4.
- * Padding - variable-length field. It MUST be zeroed on transmit and ignored on receipt.

Padding TLV MAY be used to generate IntoAM Control messages of different lengths. That capability is necessary to perform PMTUD, PMTUM, and measure synthetic packet loss and/or packet delay. When Padding TLV is used in combination with one of the performance measurement messages carried in Performance Metric TLVs as defined in Section 5.4, Padding TLV MUST follow the Performance Metric TLV.

Padding TLV MAY be used in PMTUM as part of periodically sent IntOAM Control messages. In this case, the number of consecutive messages that include Padding TLV MUST be not lesser than Detect Multiplier to ensure that the remote IntOAM peer will detect loss of messages with the Padding TLV. Also, Padding TLV MAY be present in an IntOAM Control message with the Poll flag set. If the remote IntOAM peer that supports this specification receives an IntOAM Control message with Padding TLV, it MUST include the Padding TLV with the Padding field of the same length as in the received packet and set the Final flag.

5.3. Diagnostic TLV

Diagnostic TLV MAY be used to characterize the result of the last requested operation.

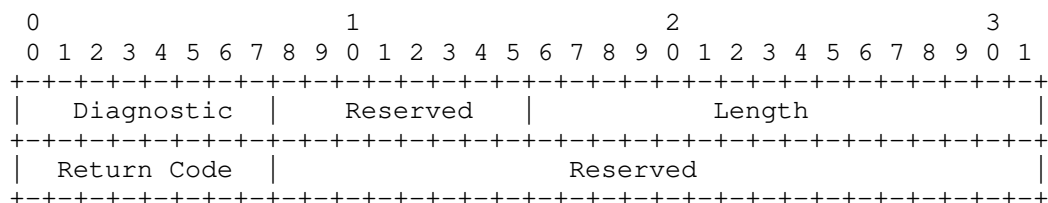


Figure 5: Diagnostic TLV Format

where fields are defined as the following:

- * Diagnostic - one-octet field. Its value (TBA6) allocated by IANA in Section 6.
- * Reserved - one-octet field. MUST be zeroed on the transmit and ignored on the receipt.
- * Length - tw-octet field. Its value MUST be set to eight.
- * Return Code - eight-bit field. The responding IntOAM system can set it to one of the values defined in Section 6.3.
- * Reserved - 24 bits-long field. MUST be zeroed on transmit and ignored on receipt.

5.4. Performance Measurement with IntOAM Control Message

Loss measurement, delay measurement, and loss/delay measurement messages can be used in the IntOAM Control message to obtain respective one-way and round-trip metrics. All the messages are encapsulated as TLVs with Type values allocated by IANA, Section 6.

The sender MAY use the Performance Metric TLV (presented in Figure 6) to measure performance metrics and obtain the measurement report from the receiver.

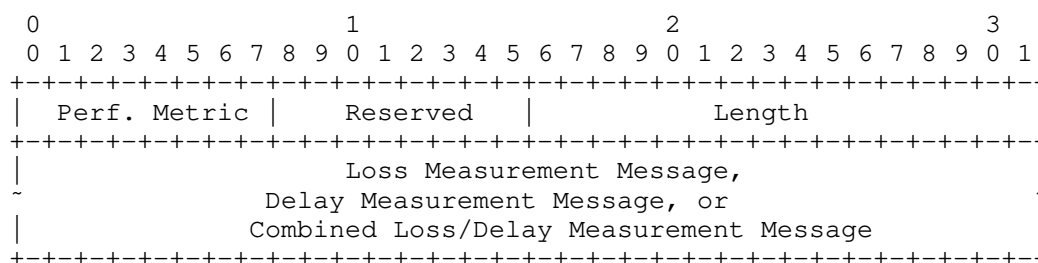


Figure 6: Performance Metric TLV Format

where fields are defined as the following:

- * Performance Metric - one-octet field. Valid values are TBA3 through TBA5 allocated by IANA in Section 6 as follows:
 - TBA3 - Loss Measurement Type;
 - TBA4 - Delay Measurement Type;
 - TBA5 - Combined Loss/Delay Measurement Type
- * Reserved - one-octet field. MUST be zeroed on the transmit and ignored on the receipt.
- * Length - two-octet field equals length on the Performance Metric TLV in octets. The value of the Length field MUST be a multiple of 4.
- * Value - various performance metrics measured either directly or using synthetic methods accordingly using the messages defined in Sections 3.1 through 3.3 [RFC6374].

To perform one-way loss and/or delay measurement, the IntoAM node MAY periodically transmit the IntoAM message with one of the TLVs listed above in Asynchronous mode. To perform synthetic loss measurement, the sender MUST monotonically increment the counter of transmitted test packets. When using Performance Metric TLV for synthetic measurement, an IntoAM Control message MAY also include Padding TLV. In that case, the Padding TLV MUST immediately follow Performance Metric TLV. Also, direct-mode loss measurement, as described in [RFC6374], is supported. Procedures to negotiate and manipulate transmission intervals defined in Sections 6.8.2 and 6.8.3 in [RFC5880] SHOULD be used to control the performance impact of using the IntoAM for performance measurement in the particular IntoAM session.

To measure the round-trip loss and/or delay metrics, an IntoAM node transmits the IntoAM Control message with the Performance Metric TLV with the Poll flag set. Before transmitting the IntoAM Control message with the Performance Metric TLV, the receiver MUST clear the Poll flag and set the Final flag.

5.5. Lightweight Authentication

Using IntoAM without any security measures, such as exchanging IntoAM Control messages without authentication, increases the risk of an attack, especially over multiple nodes. Thus, using IntoAM without security measures may cause false positive or false negative defect detection situations. In the former, an attacker may spoof IntoAM Control messages pretending to be a remote peer and can thus view the IntoAM session operation even though the real path had failed. In the latter, the attacker may spoof an altered IntoAM control message indicating that the IntoAM session is un-operational even though the path and the remote IntoAM peer operate normally.

BFD [RFC5880] allows for optional authentication protection of BFD Control messages to minimize the chances of attacks in a networking system. However, at least some of the supported authentication protocols do not provide sufficient protection in modern networks. Also, the current BFD technology requires authentication of each BFD Control message. Such an authentication requirement can put a computational burden on networking devices, especially in the Asynchronous mode, at least because authenticating each BFD Control message can require substantial computing resources to support packet exchange at high rates.

This specification defines a lightweight on-demand mode of authentication for an IntoAM session. The lightweight authentication is an optional mode. The mechanism includes negotiation (Section 5.5.1) and on-demand authentication (Section 5.5.2) phases.

During the former, IntoOAM peers advertise supported authentication capabilities and independently select the commonly supported mode of the authentication. In the authentication phase, each IntoOAM system transmits, at certain events or periodically, authenticated IntoOAM Control messages in Poll Sequence.

5.5.1. Lightweight Authentication Mode Negotiation

Figure 7 displays the format of the Authentication field that is part of the Capability Encoding:

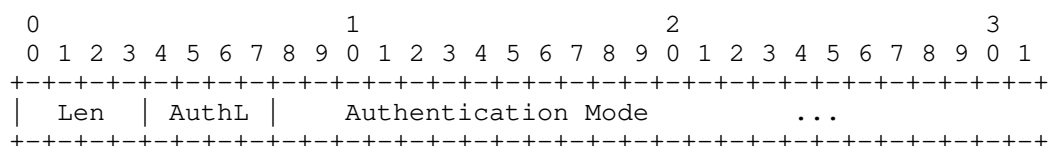


Figure 7: Lightweight Authentication Capability Field Format

where fields are defined as the following:

- * Len (Length) - four-bit field. The value of the Length field is equal to the length of the Authentication field, including the Length, in octets.
- * AuthL (Authentication Length) - four-bit field. The value of the field is, in four octets long words, the longest authentication signature the IntoOAM system is capable of supporting for any of the methods advertised in the AuthMode field.
- * Authentication Mode - variable-length field. It is a bit-coded field that an IntoOAM system uses to list modes of lightweight authentication it supports.

An IntoOAM system uses Capability TLV, defined in Section 5.1, to discover the commonly supported mode of the Lightweight Authentication. The system sets the values in the Authentication field according to properly reflect its authentication capabilities. The IntoOAM system transmits the IntoOAM Control message with Capability TLV as the first in a Poll Sequence. The remote IntoOAM system that supports this specification receives the IntoOAM Control message with the advertised Lightweight Authentication modes and stores information locally. The system responds with the advertisement of its Lightweight Authentication capabilities in the IntoOAM Control message with the Final flag set. Each IntoOAM system uses local and received information about Lightweight Authentication capabilities to deduce the commonly supported modes and selects from

that set to use the strongest authentication with the longest signature. If the common set is empty, i.e., none of supported by one IntoAM system authentication method is supported by another, an implementation MUST reflect this in its operational state and SHOULD notify an operator.

5.5.2. Using Lightweight Authentication

After IntoAM peers select an authentication mode for use in Lightweight Authentication each IntoAM system MUST use it to authenticate each IntoAM Control message transmitted as part of a Poll Sequence using Lightweight Authentication TLV presented in Figure 8.

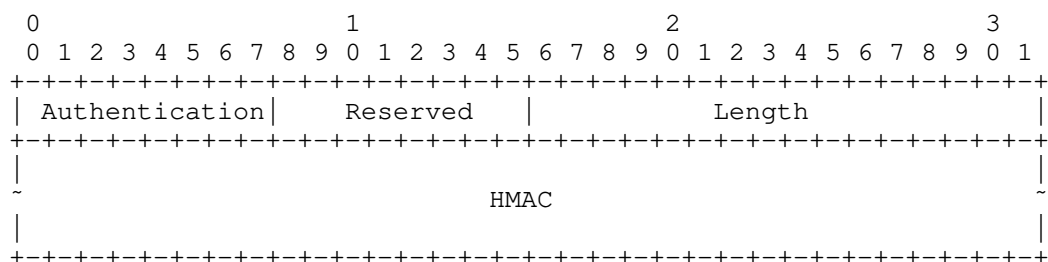


Figure 8: Lightweight Authentication TLV Format

where fields are defined as the following:

- * Lightweight Authentication - one-octet field. Its value (TBA8) allocated by IANA in Section 6
- * Reserved - one-octet field. MUST be zeroed on the transmit and ignored on the receipt.
- * Length - two-octet long field. The value equals length on the Lightweight Authentication TLV field in octets. The value of the Length field MUST be a multiple of 4.
- * HMAC (Hashed Message Authentication Code) - variable-length field. The value is the hash value calculated on the entire preceding IntoAM Control message data.

The Lightweight Authentication TLV MUST be the last in an IntoAM Control message. Padding TLV (Section 5.2) MAY be used to align the length of the IntoAM Control message, excluding the Lightweight Authentication TLV, at multiple of 16 boundary.

The IntoOAM system that receives the IntoOAM Control message with the Lightweight Authentication TLV MUST first validate the authentication by calculating the hash over the IntoOAM Control message. If the validation succeeds, the receiver MUST transmit the IntoOAM Control message with the Final flag set and the value of the Return code field in Diagnostic TLV set to None value (Table 5). If the validation of the lightweight authentication fails, then the IntoOAM system MUST transmit the IntoOAM Control message with the Final flag set and the value of the Return Code field of the Diagnostic TLV set to Lightweight Authentication failed value (Table 5). The IntoOAM system MUST have a control policy that defines actions when the system receives the Lightweight Authentication failed return code.

6. IANA Considerations

6.1. IntoOAM Message Types

IANA is requested to create the IntoOAM TLV Type registry. All code points in the range 1 through 175 in this registry shall be allocated according to the "IETF Review" procedure specified in [RFC8126]. Code points in the range 176 through 239 in this registry shall be allocated according to the "First Come First Served" procedure specified in [RFC8126]. The remaining code points are allocated according to Table 1:

Value	Description	Reference
0	Reserved	This document
1- 175	Unassigned	This document
176 - 239	Unassigned	This document
240 - 251	Experimental	This document
252 - 254	Private Use	This document
255	Reserved	This document

Table 1: IntoOAM Type Registry

This document defines the following new values in IntoOAM Type registry:

Value	Description	Reference
TBA0	Multiple TLVs Used	This document
TBA1	Padding	This document
TBA2	Capability	This document
TBA3	Loss Measurement	This document
TBA4	Delay Measurement	This document
TBA5	Combined Loss/Delay Measurement	This document
TBA6	Diagnostic	This document
TBA8	Lightweight Authentication	This document

Table 2: IntoAM Types

6.2. Lightweight Authentication Modes

IANA is requested to create a Lightweight Authentication Modes registry. All code points in this registry shall be allocated according to the "IETF Review" procedure as specified in [RFC8126].

This document defines the following new values in the Lightweight Authentication Modes registry:

Bit Position	Value	Description	Reference
0	0x1	Keyed SHA-1	This document
1	0x2	Meticulous Keyed SHA-1	This document
2	0x4	SHA-256	This document

Table 3: Lightweight Authentication Modes

6.3. Return Codes

IANA is requested to create the IntoAM Return Codes registry. All code points in the range 1 through 250 in this registry shall be allocated according to the "IETF Review" procedure as specified in [RFC8126]. The remaining code points are allocated according to Table 4:

Value	Description	Reference
0	Reserved	This document
1- 250	Unassigned	IETF Review
251-253	Experimental	This document
254	Private Use	This document
255	Reserved	This document

Table 4: IntoAM Return Codes Registry

This document defines the following new values in IntoAM Return Codes registry:

Value	Description	Reference
0	None	This document
1	One or more TLVs was not understood	This document
2	Lightweight Authentication failed	This document

Table 5: IntoAM Return Codes

7. Security Considerations

The same security considerations as those described in [RFC5880], [RFC6374], and [RFC8562]. apply to this document. Additionally, implementations that use distribution of discriminators over the control or management plane MUST use secure channels to protect systems from an infinite number of IntoAM sessions being created.

In some environments, an IntoOAM session can be instantiated using a bootstrapping mechanism supported by the control or management plane. As a result, the three-way handshaking mechanism between IntoOAM systems is bypassed. That could cause the situation where one of the systems uses overaggressive transmission intervals that are not acceptable to the remote IntoOAM system. As a result, IntoOAM Control messages could be dropped, and the remote IntoOAM system concludes the IntoOAM session failed. The environment that does not use the three-way handshake mechanism to instantiate an IntoOAM session MUST support means to balance resources used by the IntoOAM.

8. Acknowledgements

TBD

9. References

9.1. 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>>.
- [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, DOI 10.17487/RFC6374, September 2011, <<https://www.rfc-editor.org/info/rfc6374>>.
- [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>>.
- [RFC8562] Katz, D., Ward, D., Pallagatti, S., Ed., and G. Mirsky, Ed., "Bidirectional Forwarding Detection (BFD) for Multipoint Networks", RFC 8562, DOI 10.17487/RFC8562, April 2019, <<https://www.rfc-editor.org/info/rfc8562>>.

9.2. Informative References

[RFC8655] Finn, N., Thubert, P., Varga, B., and J. Farkas,
"Deterministic Networking Architecture", RFC 8655,
DOI 10.17487/RFC8655, October 2019,
<<https://www.rfc-editor.org/info/rfc8655>>.

Authors' Addresses

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com, gregory.mirsky@ztetx.com

Xiao Min
ZTE Corp.

Email: xiao.min2@zte.com.cn

Gyan Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

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.

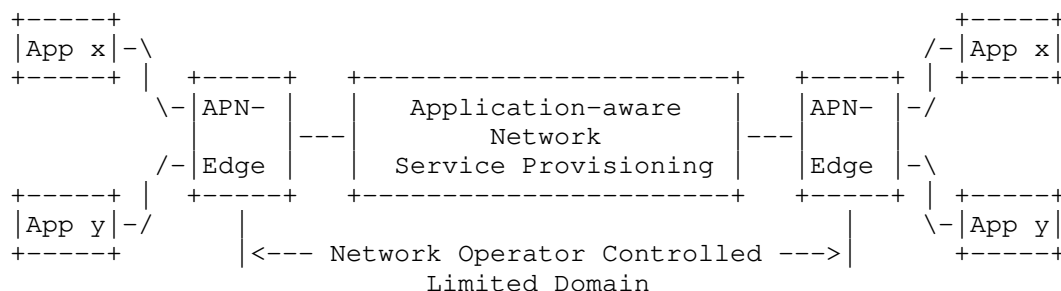


Figure 1. APN Framework and Scope

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.

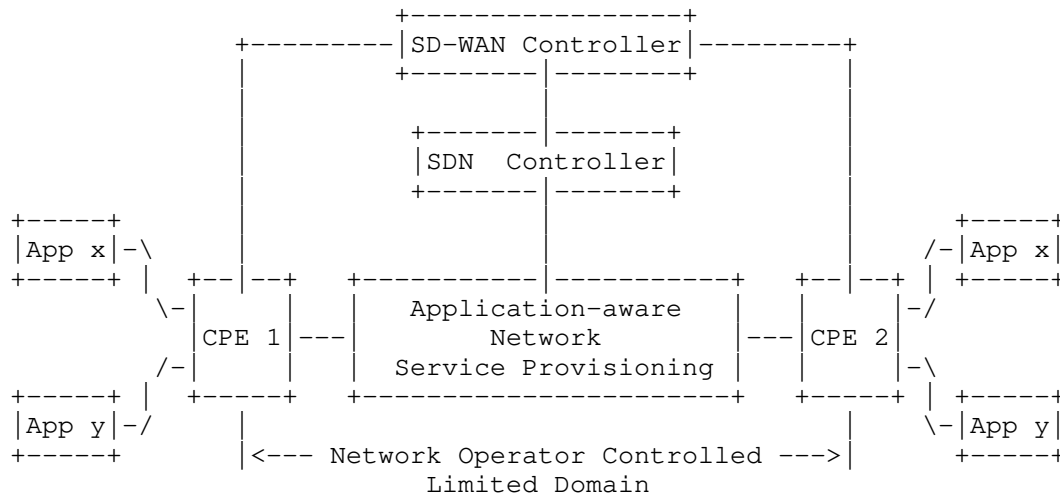


Figure 2. SD-WAN using the APN Framework

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.

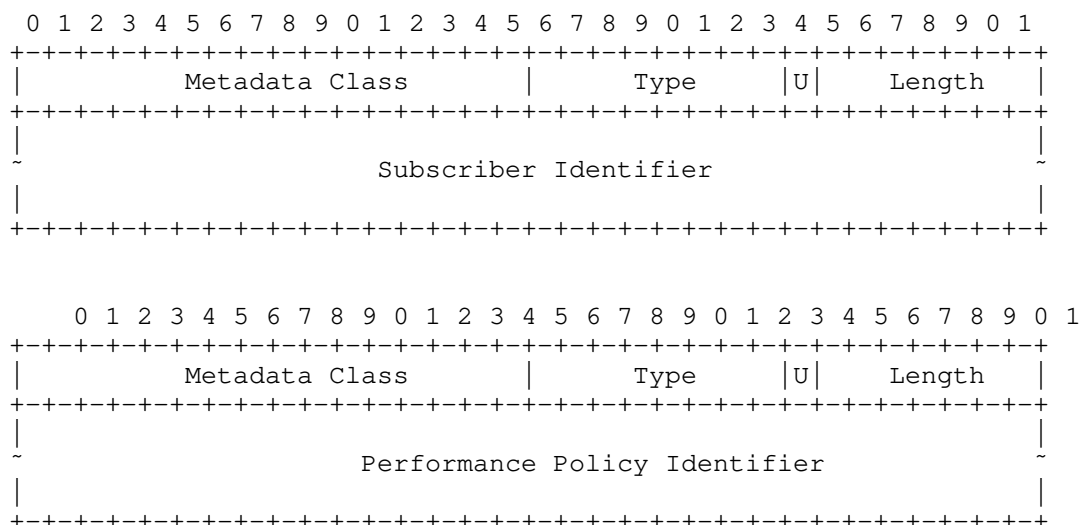


Figure 3. Subscriber Identifier and Performance Policy Identifier

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

RTGWG Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 26, 2021

F. Yang
M. Chen
T. Zhou
Huawei Technologies
February 22, 2021

Associated Channel over IPv6
draft-yang-rtgwg-ipv6-associated-channel-00

Abstract

In this document, an associated channel is introduced to provide a control channel based on IPv6, carrying types of control and management messages. By using the associated channel, messages can be transmitted between the network nodes to provide functions like path identification, OAM, protection switchover signaling, etc., targeting to provide high quality SLA guarantee to service.

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 August 26, 2021.

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. Terminology	3
3. Associated Channel	3
3.1. Identification of Associated Channel	3
3.2. ACH TLV to Carry Message	3
3.3. Encapsulation of ACH TLV in IPv6	4
3.3.1. Encapsulated in IPv6 Destination Options Header . . .	5
3.3.2. Encapsulated in IPv6 Hop-by-Hop Options Header . . .	5
3.3.3. Encapsulated in IPv6 Segment Routing Header	6
3.3.4. Encapsulated in Payload	6
3.4. Processing of ACH TLV	7
4. Applicability	7
4.1. Path Identification	7
4.2. OAM	7
4.3. Assist to Protection Switchover	7
5. IANA Considerations	8
6. Security Considerations	8
7. Acknowledgements	8
8. References	8
8.1. Normative References	8
8.2. Informative References	8
Authors' Addresses	9

1. Introduction

IPv6 is becoming widely accepted to provide the connectivity in many new emerging scenarios, including Cloud Network convergence, Cloud-Cloud interconnection, 5G vertical industries, Internet of Things, as well as the legacy networks migrating towards SR over IPv6. However, IP packet is locally lookup, and forwarded hop by hop without aware of the forwarding path. Path segment over SRv6 [I-D.ietf-spring-srv6-path-segment] provides a good solution to identify an SR path over IPv6, but can only be applicable in source routing paradigm.

To identify an IPv6 forwarding path, further to better control and manage the path, this document introduces an associated channel based on IPv6, intending to create a control channel for the control and management usages. By using the associated channel, messages can be transmitted between the network nodes to provide functions like path identification, OAM, protection switchover signaling, etc., targeting to provide high quality SLA guarantee to service.

This document also defines a TLV format for the associated channel and how it can be encapsulated in IPv6 packet, and the potential applicability in IPv6 networks. Applications of associated channel in IPv6 shall be specified in different documents and thus are out of scope of this document.

2. Terminology

OAM: Operations, Administration, and Maintenance

SLA: Service Level Agreement

ACH: Associated CHannel

3. Associated Channel

An associated channel provides a control channel that carries at least one or more types of control and management messages. The type of message is not limited to any specific usage. The associated channel is specified by two parts of information, including the identification of associated channel and the carried message.

3.1. Identification of Associated Channel

The identification of associated channel indicates the path where the packets of associated channel are transmitted on. This identification also indicates the same path of the service forwarding path which the associated channel is associated to.

3.2. ACH TLV to Carry Message

An Associated CHannel (ACH) TLV is designed to carry the message of an associated channel. ACH TLV has the following format:

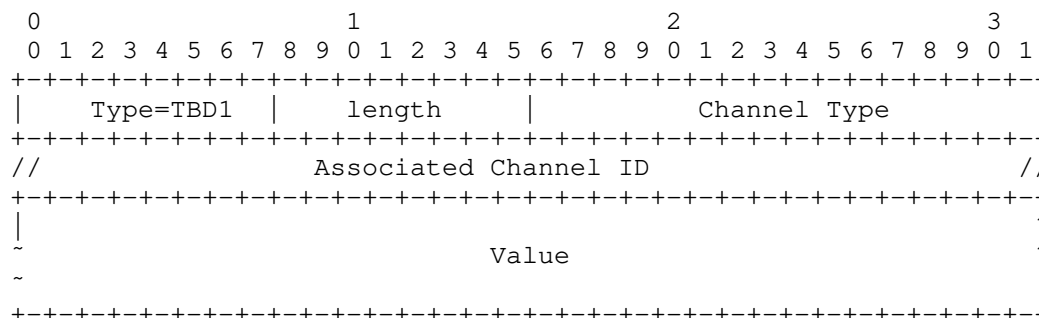


Figure 1 ACH TLV Format

Type: 8 bits, indicates it is an associated channel (ACH) TLV, and request a value assigned by IANA. The uniform type of TLV generalizes the applicability of ACH TLV to support various types of messages.

Length: 8 bits, defines the length of Value field in bytes.

Channel Type: is a 16-bit-length fixed portion as a part of Value field. It indicates the specific type of messages carried in associated channel. Note that a new ACH TLV Channel Type Registry would be requested to IANA. In the later documents which specify application protocols of associated channel, MUST also specify the applicable Channel Type field value assigned by IANA.

Associated Channel ID: indicates the identification of associated channel. The length is TBD.

Value: is a variable part of Value field. It specifies the messages indicated by Channel Type and carried in associated channel. Note that the Value field of ACH TLV MAY contain sub-TLVs to provide additional context information to ACH TLV.

3.3. Encapsulation of ACH TLV in IPv6

In the context of IPv6, ACH TLV can be encapsulated in different types of IPv6 extension header or even IPv6 payload. Note that, no matter which way ACH TLV is applied, there is no semantic change to IPv6 extension headers. Moreover, ACH TLV can be carried either with user data in an in-situ way, or in a independent synthetic packet.

3.3.1. Encapsulated in IPv6 Destination Options Header

ACH TLV can be encapsulated in IPv6 Destination Options Header as the TLV-encoded options. Figure 2 gives an example of an ACH TLV encapsulated in IPv6 Destination Options Header.

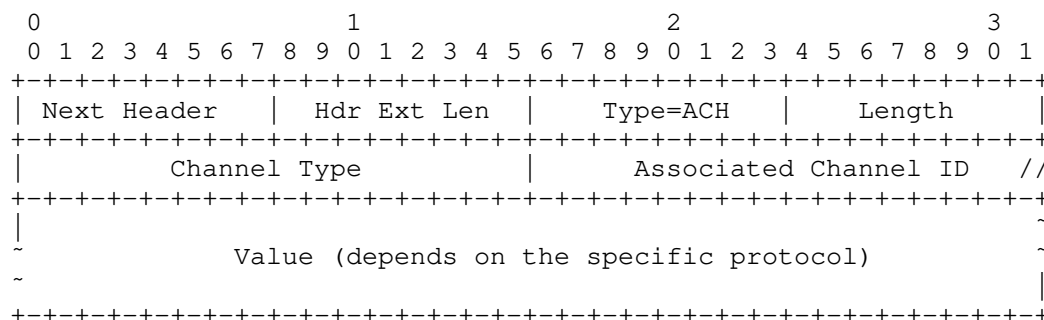


Figure 2 ACH TLV in IPv6 Destination Options Header

According to the note 1 and note 3 described in section 4.1 of[RFC8200], ACH TLV encapsulated in IPv6 Destination Options Header can provide two semantics of associated channel. When only IPv6 Destination Options Header exists or IPv6 Destination Options Header exists after the Routing Header, an end to end associated channel is provided to transmit the messages between two endpoints. When both IPv6 Destination Options Header and Routing Header exist, and IPv6 Destination Options Header exists before the Routing Header, an associated channel is provided at network nodes of the first destination that appears in the IPv6 Destination Address field plus subsequent destinations listed in the Routing header.

3.3.2. Encapsulated in IPv6 Hop-by-Hop Options Header

ACH TLV can be encapsulated in IPv6 Hop-by-Hop Options Header as the TLV-encoded options. Same option type numbering space is used for both Hop-by-Hop Options header and Destination Options header. Similarly, the ACH TLV in IPv6 Hop-by-Hop Options Header shares the same encapsulation shown in Figure 2.

When it is encapsulated in IPv6 Hop-by-Hop Options Header, it provides an associated channel at every node along the forwarding path.

3.3.3. Encapsulated in IPv6 Segment Routing Header

ACH TLV can be encapsulated in IPv6 Segment Routing Header, as SRH optional TLV. Figure 3 gives an example of an ACH TLV encapsulated in IPv6 Segment Routing Header.

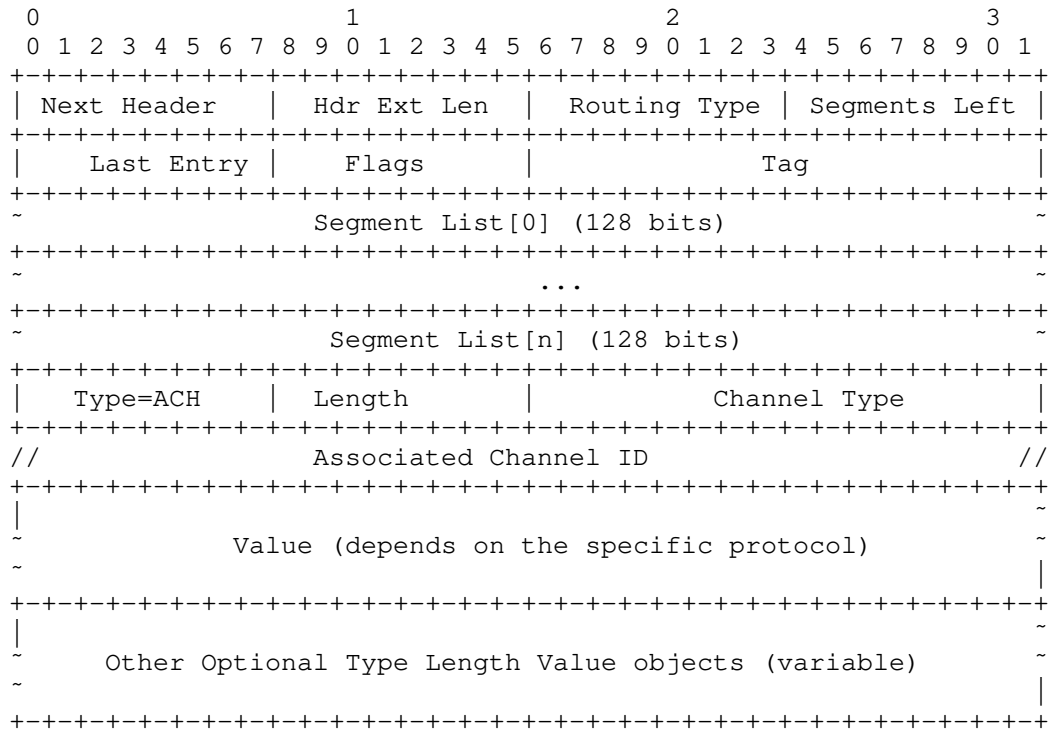


Figure 3 ACH TLV in IPv6 Segment Routing Header

When ACH TLV is encapsulated in IPv6 Segment Routing Header, it provides an associated channel at every SRv6 endpoints along the path.

3.3.4. Encapsulated in Payload

ACH TLV can also be encapsulated in the payload of an IPv6 packet. The term of payload here means the octets after the IPv6 header and extension headers. A synthetic packet is created with the payload of messages. and transmitted in an associated channel. The synthetic packet can use the same routing information with service data whose associated channel is associated to. For example, synthetic packet can encapsulate the same segment list as the one used in IPv6 SRH of service data.

3.4. Processing of ACH TLV

Take the ACH TLV encapsulated in Segment Routing Header as an example. At headend, ACH TLV is encapsulated with control and management messages in Segment Routing Header. When midpoint or tail-end receives an SRv6 packet with ACH TLV, it recognizes the ACH TLV, check the Channel Type field to interpret the protocol, and continue with processing of messages. The processing of message is not limited, for example READ or/and WRITE. It should depend on the specification of protocols used in the associated channel.

4. Applicability

4.1. Path Identification

In a native IPv6 network, packets is transmitted hop by hop, there is no way to identify an IPv6 forwarding path. The path needs to be identified when OAM or protection switchover is applied to the path.

4.2. OAM

OAM includes the a group of functions such as connectivity verification, fault indication and detection, and performance measurement of loss and delay etc. For example, BFD defines a generic control packet format that can be encapsulated in different data planes to provide low-overhead and short-duration failure detection function. The format can also be encapsulated in ACH TLV as the option TLV of Destination Options Header, to provide the same connectivity verification and fault detect functions without introducing upper layer protocols. Another example is to encapsulate PDU formats of Ethernet OAM [ITU-T G.8013] in Value field of ACH TLV to provide a set of OAM functions. By using ACH TLV to carry OAM messages in associated channel, different OAM functions can be easily integrated. The OAM functions can be performed in either end-to-end or hop-by-hop mode. For example, signal degrade happens on the intermediate node could be discovered and further indicated in associated channel to monitor the path status.

4.3. Assist to Protection Switchover

Linear protection [RFC6378] provides a very flexible protection mechanism in a mesh network because it can operate between any pair of endpoints. ACH TLV can be used to transmit the protection state control messages on an IPv6 forwarding path to provide the function of bidirectional protection switchover.

5. IANA Considerations

- o This document requests IANA to assign a codepoint of Destination Options and Hop-by-Hop Options.
- o This document requests IANA to assign a codepoint of Segment Routing Header TLVs to indicate ACH TLV.
- o This document request IANA to create a new IANA-managed registry of ACH Channel Type to identify the usage of associated channel.

6. Security Considerations

TBD

7. Acknowledgements

TBD

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>>.

8.2. Informative References

- [I-D.ietf-spring-srv6-path-segment] Li, C., Cheng, W., Chen, M., Dhody, D., and R. Gandhi, "Path Segment for SRv6 (Segment Routing in IPv6)", draft-ietf-spring-srv6-path-segment-00 (work in progress), November 2020.
- [RFC6378] Weingarten, Y., Ed., Bryant, S., Osborne, E., Sprecher, N., and A. Fulignoli, Ed., "MPLS Transport Profile (MPLS-TP) Linear Protection", RFC 6378, DOI 10.17487/RFC6378, October 2011, <<https://www.rfc-editor.org/info/rfc6378>>.
- [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

Fan Yang
Huawei Technologies
Beijing
China

Email: shirley.yangfan@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Beijing
China

Email: mach.chen@huawei.com

Tianran Zhou
Huawei Technologies
Beijing
China

Email: zhoutianran@huawei.com