Network Working Group
Internet Draft
Intended status: Experimental
Expires: December 27, 2019

L. Dunbar
Futurewei
S. Hares
Hickory Hill Consulting

June 27, 2019

Subsequent Address Family Indicator for SDWAN Ports
draft-dunbar-idr-sdwan-port-safi-03

Abstract

The document specifies a new BGP NLRI and SAFI for advertising WAN ports properties of a SDWAN edge node. SDWAN edge node's WAN ports may face untrusted networks, such as the public internet, may get assigned IP addresses from the Internet Service Providers (ISPs), may get assigned dynamic IP addresses via DHCP, or may have private addresses (e.g. inside third party Cloud DCs). Packets forwarded through those SDWAN WAN ports might need to be encrypted (depending on the user policies) or need to go through NAT. SDWAN edge nodes need to propagate those WAN ports properties to the peers who are authorized to communicate across different types of underlay networks including the untrusted networks.

BGP Route Reflectors (RR) are proposed as the entities to propagate the WAN ports properties of SDWAN edge nodes to a controlled group of edges reachable via overlay networks.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on December 27, 2019.

Copyright Notice

Table of Contents

1. Introduction

[Net2Cloud-Problem] introduces using SDWAN to reach dynamic workloads in multiple third-party data centers and aggregate multiple underlay paths, including public untrusted networks, provided by different service providers.

[SDWAN-BGP-USAGE] describes multiple SDWAN scenarios and how/why using BGP as control plane for the SDWAN networks.

[Tunnel-Encap] describes how to construct a BGP UPDATE messages that advertise endpoints' tunnel encapsulation capabilities and the respective attached client routes, so that the receivers of the BGP UPDATE can establish appropriate tunnels to the endpoints for the aforementioined client routes. [Tunnel-Encap] has a "Remote endpoint subTLV" for one node to advertise another node's encapsulation capabilities. The receivers of the Tunnel UPDATE would construct the encapsulation header with the Outer Destination Address equal to the address carried in the "Remote Endpoint sub-TLV". All those have nothing to do with the SDWAN Edge WAN ports properties registration.

This document describes a new BGP NLRI and SAFI for SDWAN edge nodes to register (or propagate) their WAN ports properties. This new SAFI & NLRI is for a scenario where one SDWAN edge node has multiple WAN ports, some of which connected to private networks and others connected to public untrusted networks [Scenario #2 described in the [SDWAN-BGP-USAGE]]. The same routes attached to the SDWAN can be sent/received over the private networks without encryption (for better performance) and sent/received over the public networks with encryption.

The [SDWAN-BGP-USAGE] document describes the following functional components of the control plane for the scenario (i.e. the SDWAN Scenario #2):

. Each Edge SDWAN edge node is informed of its SDWAN controller, either burned in the node or configured, which is the BGP RR in this document.
. Each SDWAN edge node needs to advertise its WAN ports properties via the secure channel with the RR. RR then propagates the received WAN ports properties to the authorized peers based on appropriate policies. Because the connection among SDWAN edges and the RR can be public untrusted networks, the communication session between RR and SDWAN edges MUST run over a secure channel (e.g. TLS, DTLS, or others).

. SDWAN edges pairwise secure channel establishment, such as IPsec parameters negotiation, public key exchange, etc, and

. Client routes distribution, just like EVPN or L3VPN using [Tunnel-Encap] to advertise all possible tunnels for clients routes.

The new SDWAN NLRI and SAFI can also include information such as WAN port's NAT information, SDWAN-SITE-ID, SDWAN EdgeNode-ID, and IPsec related information.

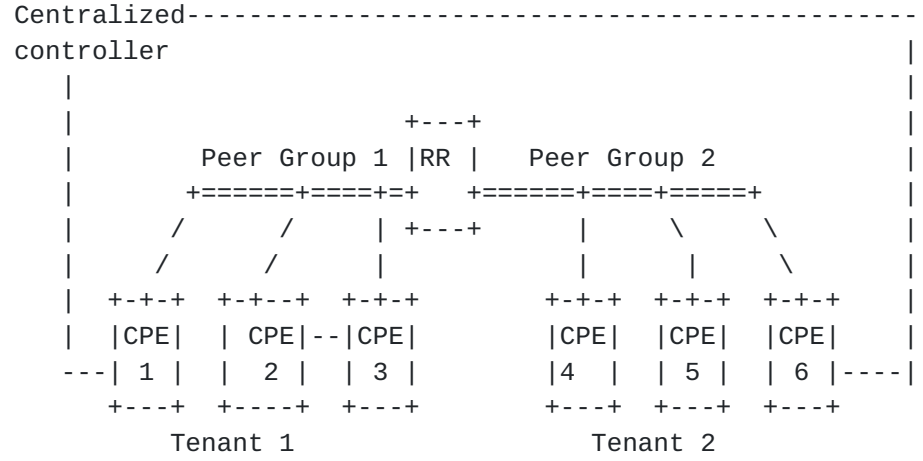


Figure 1: SDWAN WAN Port Properties Advertisement via RR

Note: All CPEs (CPE1, CPE2, CPE, CPE4, CPE5, and CPE) connect to the centralized controller, but only 2 connections are show in this diagram.

2. Conventions used in this document

Cloud DC: Off-Premise Data Centers that usually host applications and workload owned by different organizations or tenants.

Controller: Used interchangeably with SDWAN controller to manage SDWAN overlay path creation/deletion and monitor the path conditions between sites.

CPE-Based VPN: Virtual Private Secure network formed among CPEs. This is to differentiate from most commonly used PE-based VPNs a la RFC 4364.

SDWAN End-point: An WAN port (logical or physical) of a SDWAN edge node. (If "endpoint" is used, it refers to a SDWAN End-point).

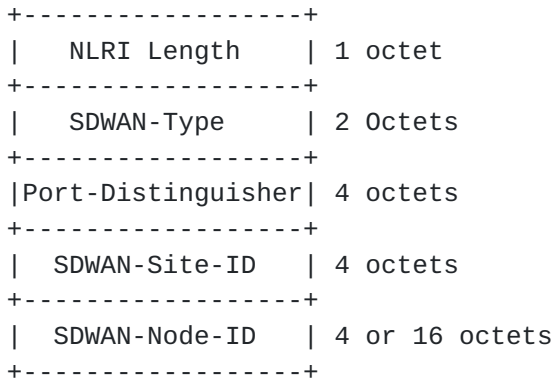
OnPrem: On Premises data centers and branch offices

SDWAN: Software Defined Wide Area Network. In this document, "SDWAN" refers to the solutions of pooling WAN bandwidth from multiple underlay networks to get better WAN bandwidth management, visibility & control. When the underlay networks are private networks, traffic can be forwarded without additional encryption; when the underlay networks are public, such as Internet, some traffic needs to be encrypted when forwarding through those WAN ports(depending on user provided policies).

3. SDWAN NLRI Format

The new SAFI code point 74 has been assigned by IANA as the Subsequent Address Family Identifier for advertising properties of WAN ports that face untrusted networks. Depending on user policies, some packets sent through those WAN ports will need encryption.

The SDWAN SAFI (code point 74 assigned by IANA) uses a new NLRI defined as follows:



where:

- NLRI Length: 1 octet of length expressed in bits as defined in [RFC4760].
- SDWAN-Type: to define the encoding of the rest of the SDWAN NLRI.
- Port Distinguisher: SDWAN edge node Port identifier. There can be many ports on a SDWAN edge node; each port can have different properties. For example, some ports may get ISP or DHCP assigned IP addresses (IPv4 or IPv6), some may have private IP addresses that packets to/from those ports have to traverse NAT. The detailed properties about the port are further encoded in the subsequent subTLVs, e.g. Port-subTLV.

- SDWAN-Site-ID: used to identify a common property shared by a set of SDWAN edge nodes, such as the property of a specific geographic location shared by a group of SDWAN edge nodes.
- SDWAN EdgeNode ID: the SDWAN edge node identifier, which can be the node's system ID or the loopback address (IPv4 or IPv6) of the SDWAN edge node.

The content of the SDWAN Port properties is encoded in the Tunnel Encapsulation Attribute originally defined in [Tunnel-Encap] using a

new Tunnel-Type TLV (code point to be assigned by IANA from the "BGP Tunnel Encapsulation Attribute Tunnel Types" registry).

SDWAN SAFI (=74) NLRI: < SDWAN-Type, Length, Port-distinguisher, SDWAN-Site-ID, SDWAN-Node-ID>
Attributes:
   Tunnel Encaps Attribute
       Tunnel Type: SDWAN Port Property
           NAT SubTLV
           IPsec-SA Attribute SubTLV
           Port-subTLV

Where
    - NAT SubTLV is for describing additional information about the SDWAN tunnel end-points, such as NAT property.
    - IPsec-SA SubTLV is for the node to establish IPsec SA with other peers.
    - Port-subTLV is for additional properties of the WAN port.

The Tunnel Encaps Attribute are defined as follows:

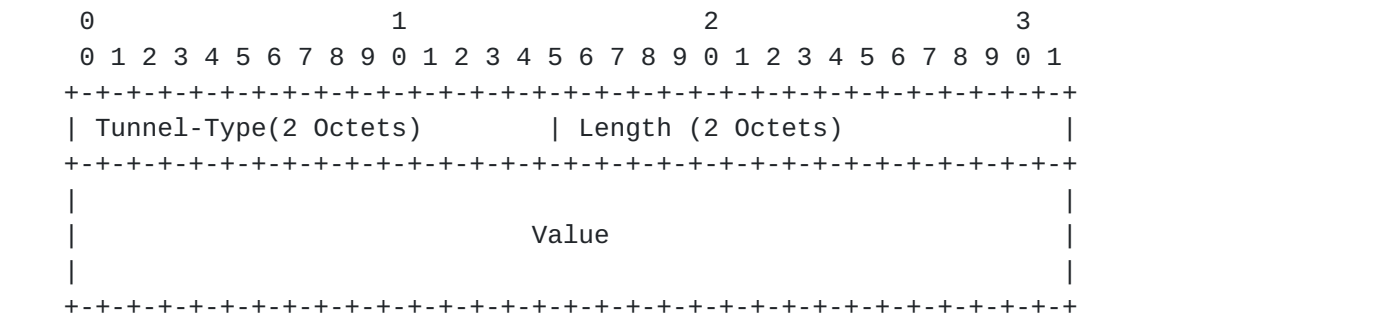


Figure 1: SDWAN Tunnel Encapsulation TLV Value Field

Where:
   Tunnel Type is SDWAN Port Property (to be assigned by IANA).

3.1. SDWAN Route Type

A new Route Type that defines the encoding of the rest of the SDWAN NLRI, and a set of sub-TLVs to specify its end-point attributes, policies associated with the Ports:

3.2. Port Distinguisher

One (SDWAN Edge) node can have multiple ports, and each port can support multiple IPsec SA to different peers. The Port Distinguisher is to uniquely identify a port (or link).

The property of the port are encoded in the subTLV attached to the SDWAN NLRI:

a) The IP address (IPv4 or IPv6) & AS number of the Port
b) NAT information for ports with Private IP address
c) IPsec Security Association related information if the port is facing public network and traffic through which have to be encrypted.

Detailed encoding for those properties is described in Section 3.4 & Section 3.5 respectively.

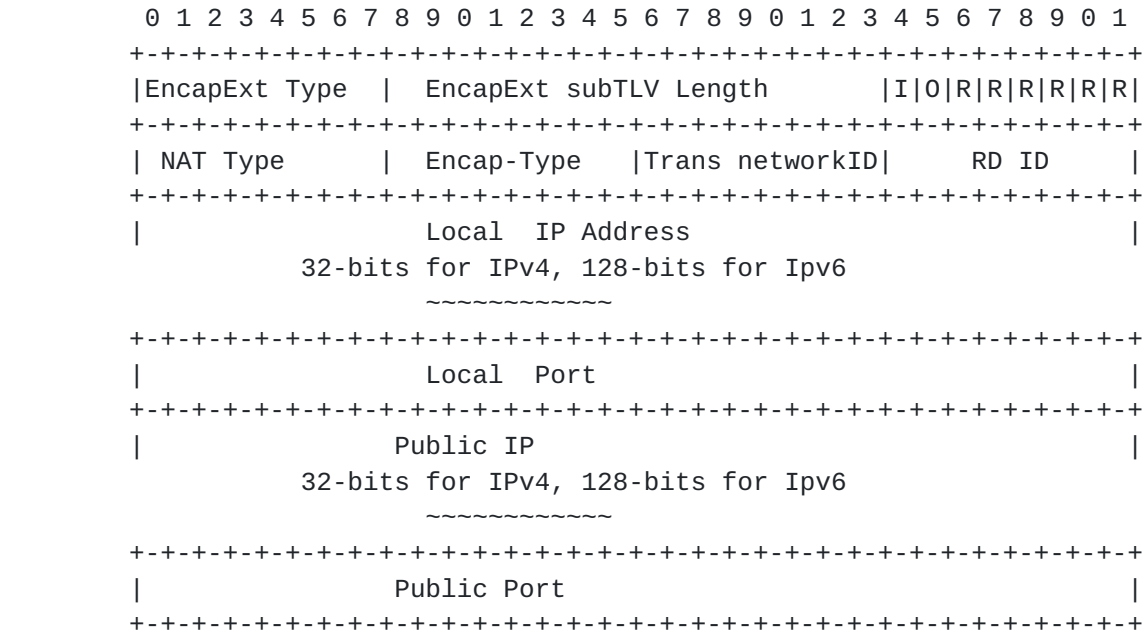
3.3. SDWAN Site ID

SDWAN Site ID is used to identify a common property shared by a set of SDWAN edge nodes/ports, such as the property of a specific geographic location. The property is used to steer an overlay route to traverse specific geographic locations for various reasons, such as to comply regulatory rules, to utilize specific value added services, or others.

3.4. Extended Port Property

EncapExt sub-TLV is for describing additional information about a SDWAN port, such as the NAT property if the port has private address, the network identifier that the port is part of, etc.

A SDWAN edge node can inquire STUN (Session Traversal of UDP Through Network Address Translation RFC 3489) Server to get the NAT property, the public IP address and the Public Port number to pass to peers.



Where:

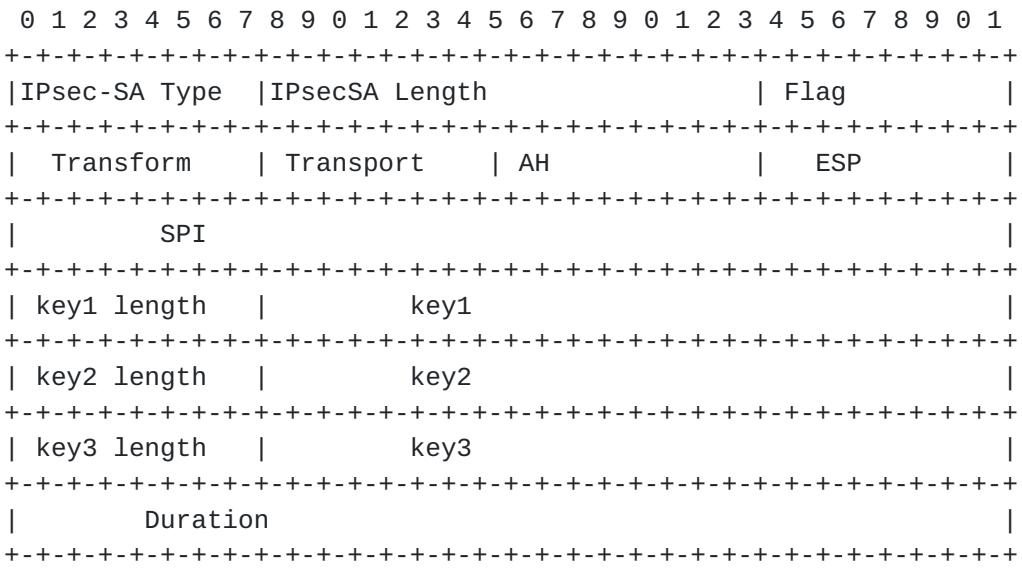
- o EncapExt Type: indicate it is the EncapExt SubTLV.
- o EncapExt subTLV Length: the length of the subTLVE.
- o Flags:
  - I bit (CPE port address or Inner address scheme)
    If set to 0, indicate the inner (private) address is IPv4.
    If set to 1, it indicates the inner address is IPv6.
  - O bit (Outer address scheme):
    If set to 0, indicate the public (outer) address is IPv4.
    If set to 1, it indicates the public (outer) address is IPv6.
  - R bits: reserved for future use. Must be set to 0 now.

- o NAT Type.without NAT; 1:1 static NAT; Full Cone; Restricted Cone; Port Restricted Cone; Symmetric; or Unknown (i.e. no response from the STUN server).

- o Encap Type.the supported encapsulation types for the port facing public network, such as IPsec+GRE, IPsec+VxLAN, IPsec without GRE, GRE (when packets don't need encryption)
- o Transport Network ID.Central Controller assign a global unique ID to each transport network.
- o RD ID.Routing Domain ID.Need to be global unique.
- o Local IP.The local (or private) IP address of the port.
- o Local Port.used by Remote SDWAN edge node for establishing IPsec to this specific port.
- o Public IP.The IP address after the NAT. If NAT is not used, this field is set to NULL.
- o Public Port.The Port after the NAT. If NAT is not used, this field is set to NULL.

3.5. IPsec Security Association Property

The IPsecSA sub-TLV is for the SDWAN edge node to establish IPsec security association with their peers via the port that face untrusted network:



Where:

- o IPsec-SA SubTLV Type: to be assigned by IANA. The type value has to be between 128~255 because IPsec-SA subTLV needs 2 bytes for length to carry the needed information.
- o IPsec-SA subTLV Length (2 Byte): 25 (or more)
- o Flags: 1 octet of flags. None are defined at this stage. Flags SHOULD be set to zero on transmission and MUST be ignored on receipt.
- o Transform (1 Byte): the value can be AH, ESP, or AH+ESP.
- o Transport (1 byte): the value can be Tunnel Mode or Transport mode
- o AH (1 byte): AH authentication algorithms supported, which can be md5 | sha1 | sha2-256 | sha2-384 | sha2-512 | sm3. Each SDWAN edge node can have multiple authentication algorithms; send to its peers to negotiate the strongest one.
- o ESP (1 byte): ESP authentication algorithms supported, which can be md5 | sha1 | sha2-256 | sha2-384 | sha2-512 | sm3. Each SDWAN edge node can have multiple authentication algorithms; send to its peers to negotiate the strongest one. Default algorithm is AES-256.
- o SPI: 4 bytes
- o Key1.AH authentication key
- o Key2.ESP authentication key
- o Key3.ESP encryption "public" key
- o Duration: SA life span.

3.6. Remote Endpoint

The Remote Endpoint sub-TLV is not used for SDWAN NLRI because

- o The SDWAN EdgeNode ID and Site ID are already encoded in the SDWAN NLRI,
- o The network connected by the SDWAN WAN port might have identifier that is more than the AS number. SDWAN controller might use its own specific identifier for the network.
- o The Transport-Network-ID in the EncapExt sub-TLV represents the SDWAN unique network identifier.

If the Remote Endpoint Sub-TLV is present, it is ignored by other SDWAN edge nodes.

4. Operation of SDWAN Edge Node:

Using Figure 1 as illustration, the processing steps to announce the SDWAN combination of routes, NAT and IPsec information via BGP are:

1) Advertise the SDWAN port properties, such as Port identifiers and supported properties etc. to RR via the SDWAN SAFI NLRI.

2) RR propagate the information to CPE2 & CPE 3.
3) CPE2 and CPE3 can choose to establish IPsec SA with the CPE1 after receiving the CPE1 WAN properties from RR.

Note: Tenant separation is achieved by peer group policies on the RR.

4. Manageability Considerations

TBD - this needs to be filled out before publishing

5. Security Considerations

The document describes the encoding for SDWAN edge nodes to advertise its SDWAN WAN ports properties to their peers via untrusted & unsecure networks.

The secure propagation is achieved by secure channels, such as TLS, SSL, or IPsec, between the SDWAN edge nodes and the local controller RR.

[More details need to be filled in here]

6. IANA Considerations

This document requires the following IANA actions.

- SDWAN Overlay SAFI = 74 assigned by IANA
- SDWAN Route Type

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

7.2. Informative References

[RFC8192] S. Hares, et al, "Interface to Network Security Functions (I2NSF) Problem Statement and Use Cases", July 2017

[RFC5521] P. Mohapatra, E. Rosen, "The BGP Encapsulation Subsequent Address Family Identifier (SAFI) and the BGP Tunnel Encapsulation Attribute", April 2009.

[Tunnel-Encap]E. Rosen, et al, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-09, Feb 2018.

[VPN-over-Internet] E. Rosen, "Provide Secure Layer L3VPNs over Public Infrastructure", draft-rosen-bess-secure-l3vpn-00, work-in-progress, July 2018

[DMVPN] Dynamic Multi-point VPN: https://www.cisco.com/c/en/us/products/security/dynamic-multipoint-vpn-dmvpn/index.html

[DSVPN] Dynamic Smart VPN: http://forum.huawei.com/enterprise/en/thread-390771-1-1.html

[ITU-T-X1036] ITU-T Recommendation X.1036, "Framework for creation, storage, distribution and enforcement of policies for network security", Nov 2007.

[Net2Cloud-Problem] L. Dunbar and A. Malis, "Seamless Interconnect Underlay to Cloud Overlay Problem Statement", draft-dm-net2cloud-problem-statement-02, June 2018

[Net2Cloud-gap] L. Dunbar, A. Malis, and C. Jacquenet, "Gap Analysis of Interconnecting Underlay with Cloud Overlay", draft-dm-net2cloud-gap-analysis-02, work-in-progress, Aug 2018.

[Tunnel-Encap] E. Rosen, et al "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-10, Aug 2018.

8. Acknowledgments

Acknowledgements to Wang Haibo, Hao Weiguo, and ShengCheng for implementation contribution; Many thanks to Jim Guichard, John Scudder, Darren Dukes, Andy Malis, Rachel Huang and Donald Eastlake for their review and contributions.

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Linda Dunbar
Futurewei
Email: ldunbar@futurewei.com

Sue Hares
Hickory Hill Consulting
Email: shares@ndzh.com