Network Working Group
Internet Draft
Intended status: Experimental
Expires: Dec 10, 2019

L. Dunbar
Futurewei
S. Hares
Hickory Hill Consulting

June 10, 2019

Subsequent Address Family Indicator for SDWAN Ports
draft-dunbar-idr-sdwan-port-safi-02

Abstract

The document specifies a new BGP NLRI and SAFI for advertising properties of a SDWAN edge node WAN ports that face untrusted networks, such as the public internet. Those WAN ports 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 sent over those SDWAN WAN ports might need to be encrypted (depending on the user policies) or need to go through NAT. SDWAN edge needs to propagate those WAN ports properties to its SDWAN controller, which propagates to the authorized peers and manage the IPsec SAs among those peers for encrypting traffic via the untrusted networks.

BGP Route Reflectors (RR) are proposed to propagate this information to the controlled group of other edges in order to allow scaling of the WAN ports property propagation in SDWAN Overlay.

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 10, 2019.

Copyright Notice

Table of Contents

1. Introduction

[Net2Cloud-Problem] introduces using SDWAN to reach workloads in dynamic third party data centers and aggregate multiple underlay paths, including public untrusted networks, provided by different service providers. However, scaling the combination of routes and IPsec SAs key management can be an issue when the number of nodes interconnected by the SDWAN overlay paths reaches 10,000 to 100,000 nodes.

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

This document describes a new BGP NLRI and SAFI to advertise properties of WAN ports facing the public internet. This new SAFI & NLRI is for the Scenario #2 of the [SDWAN-BGP-USAGE] where one "SDWAN" edge node having multiple ports some of which connected to private networks and others connected to public untrusted networks. The packets sent over the private networks can go natively without encryption (for better performance), only the packets sent over the public networks needs IPsec SA.

The new SAFI and NLRI are for advertising the properties of WAN ports facing public untrusted networks, through which data packets have to be encrypted using IPsec.

The [SDWAN-BGP-USAGE] document describes the three functional tiers for the control plane of SDWAN Scenario #2:

. Tier 1 (Edge Router of SDWAN): Each Edge SDWAN router registers with the SDWAN Controller using a secure connection (e.g. TLS). During the registration process, the controller may suggest a specific BGP RR peer for the Edge SDWAN router to exchange BGP route with.

After registering, each Edge Router sends routes + SDWAN WAN ports information (NAT and security information) via the SDWAN

SAFI + NLRI to the BGP RR. Due to the sensitivity of the information, the BGP peering session MUST be configured to run over a Secure TCP (TLS).

. Tier 2: Route Reflector that combines information from the security information from the SDWAN controller, the WAN ports properties from SDWAN edge routers, and

. Tier 3: Client routes distribution, just like EVPN or L3VPN, except including additional paths over the WAN ports facing the public Internet.

Traffic go through the private networks links natively without encryption and are encrypted when sent out the WAN ports facing public Internet.

The BGP peers use a new BGP NLRI and SAFI to pass the SDWAN Internet WAN ports properties, such as NAT and security association (SA). This information includes the Port-ID and port related NAT information, SDWAN-SITE-ID, SDWAN Node-ID, and IPsec security information.

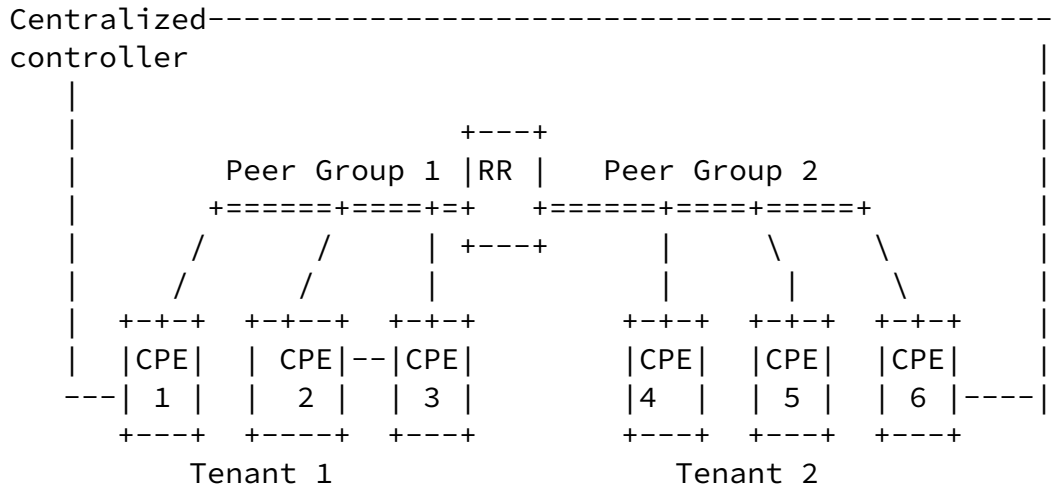


Figure 1: SDWAN Capability Advertisement via RR

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 node. (If "endpoint" is used, it refers to a SDWAN End-point).

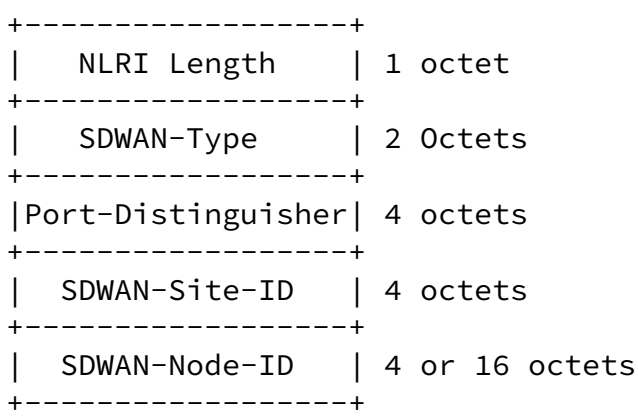
OnPrem: On Premises data centers and branch offices

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 traverse without additional encryption; when the underlay networks are public, such as Internet, some traffic needs to be encrypted when traversing through (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 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 node Port identifier. There can be many ports on a SDWAN 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 nodes, such as the property of a specific geographic location shared by a group of SDWAN nodes.
- SDWAN Node ID: the SDWAN node identifier, which can be the node's system ID or the loopback address (IPv4 or IPv6) of the SDWAN 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:

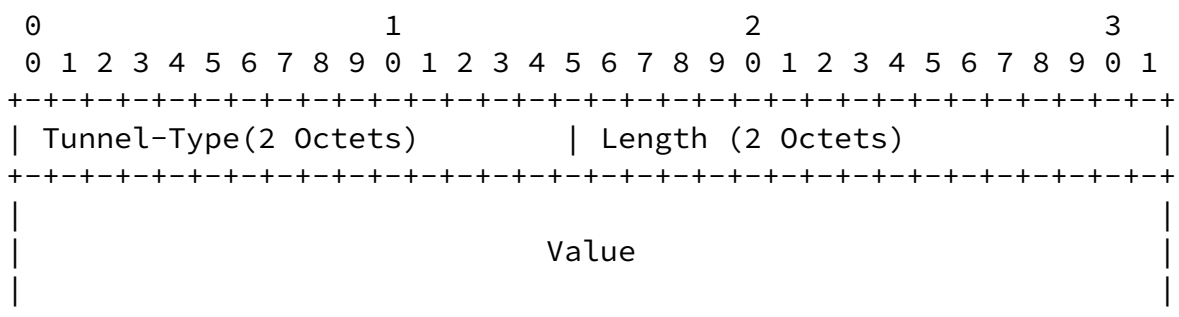


Figure 1: SDWAN Tunnel Encapsulation TLV Value Field

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) 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 are described in Section 3.4 & Section 3.5 respectively.

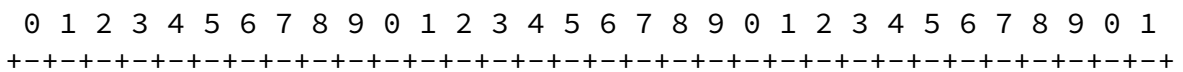
3.3. SDWAN Site ID

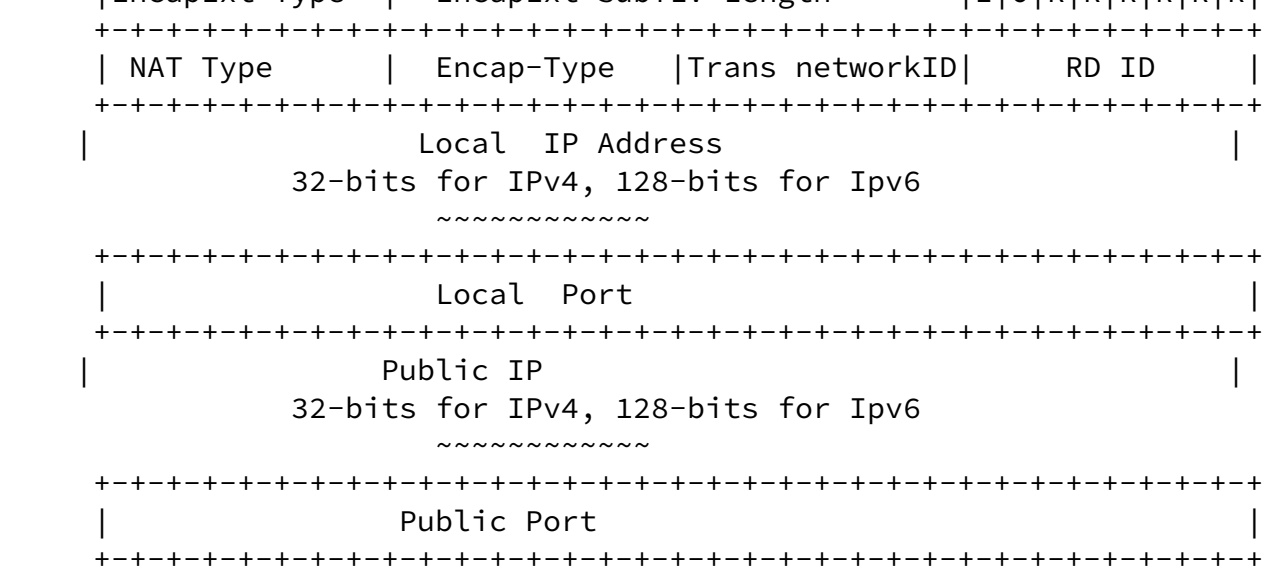
SDWAN Site ID is used to identify a common property shared by a set of SDWAN nodes/ports, such as the property of a specific geographic location. The property is used to steer an overlay route to traverse specific geographic locations for various reasons, such as to comply regulatory rules, to utilize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





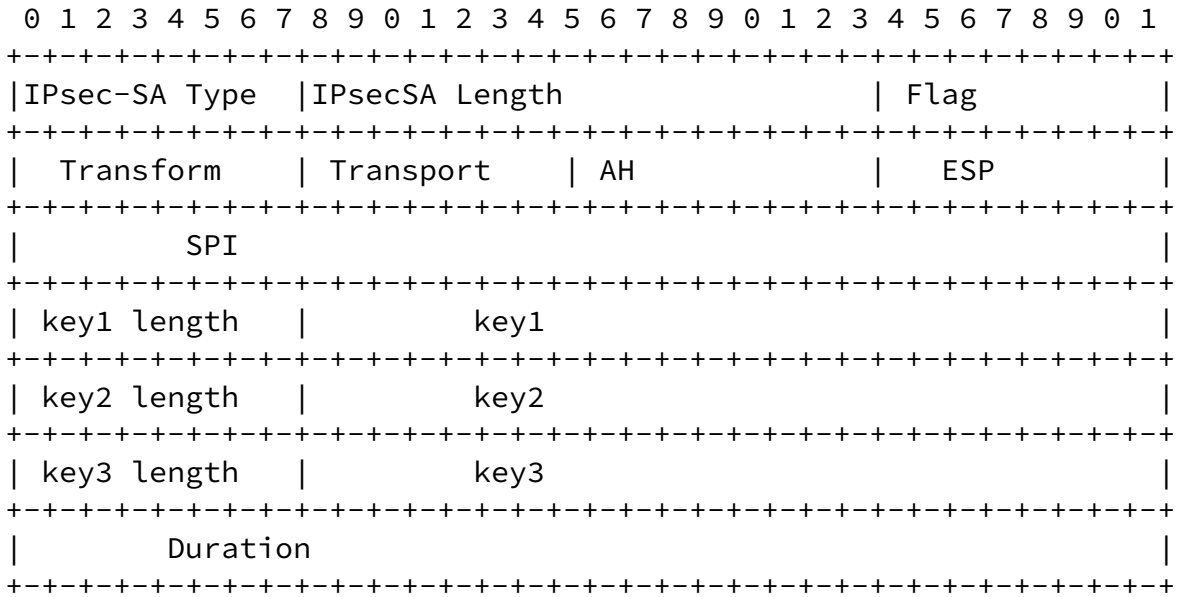
Where:

- 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 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 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 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 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 network to which a SDWAN port is connected 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 nodes.

4. Operation of SDWAN routers:

The processing steps on CPE1 to announce the SDWAN combination of routes, NAT and IPsec information via BGP are:

1. Advertise the SDWAN capability information and 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 different SDWAN nodes being added to different Peer Group.

4. Manageability Considerations

TBD - this needs to be filled out before publishing

5. Security Considerations

The document is to address how SDWAN nodes advertise its SDWAN capability to their peers via untrusted & unsecure networks.

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

[More details need to be filled in here]

6. IANA Considerations

This document requires the following IANA actions.

- o SDWAN Overlay SAFI = 74 assigned by IANA
- o 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