BGP Usage for SDWAN Overlay Networks
draft-dunbar-bess-bgp-sdwan-usage-01

Abstract

The document describes three distinct SDWAN scenarios and discusses the applicability of BGP for each of those scenarios. The goal of the document is to make it easier for future SDWAN control plane protocols discussion.

SDWAN edge nodes are commonly interconnected by multiple underlay networks that are owned and managed by different network providers. A BGP-based control plane is chosen for handling large number of SDWAN edge nodes with little manual intervention.

Status of this Memo

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

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

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 January 8, 2009.

Copyright Notice

Copyright (c) 2019 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
(http://trustee.ietf.org/license-info) in effect on the date of
publication of this document. Please review these documents
carefully, as they describe your rights and restrictions with
respect to this document. Code Components extracted from this
document must include Simplified BSD License text as described in
Section 4.e of the Trust Legal Provisions and are provided without
warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction.......................................................3
2. Conventions used in this document..............................4
3. Use Case Scenario Description and Requirements................5
   3.1. Requirements..................................................6
       3.1.1. Client Service Requirement............................6
       3.1.2. SDWAN Node Provisioning..............................6
   3.2. Scenarios #1: Homogeneous WAN............................8
   3.3. Scenario #2: SDWAN WAN ports to VPN’s PEs and to Internet.9
3.4. Scenario #3: SDWAN WAN ports to MPLS VPN and the Internet

4. Provisioning Model
   4.1. Client Service Provisioning Model
   4.2. WAN Ports Provisioning Model
      4.2.1. Why BGP as Control Plane for SDWAN WAN Ports
      Registration?

5. SDWAN Traffic Forwarding Walk Through
   5.1. SDWAN Network Startup Procedures
   5.2. Packet Walk-Through for Scenario #1
   5.3. Packet Walk-Through for Scenario #2
      5.3.1. SDWAN node WAN Ports Properties Registration
      5.3.2. Controller Facilitated IPsec SA & NAT management
      5.3.3. BGP Based SDWAN client routes
   5.4. Packet Walk-Through for Scenario #3

6. Manageability Considerations

7. Security Considerations

8. IANA Considerations

9. References
   9.1. Normative References
   9.2. Informative References

10. Acknowledgments

1. Introduction

An "SDWAN" network consists of many segments of parallel paths over different underlay networks, some of which are private networks over which traffic can traverse without encryption, others require encryption over untrusted public networks.

[Net2Cloud-Problem] describes the network related problems that enterprises face today in transitioning their IT infrastructure to support a digital economy, such as the need to connect enterprises’ branch offices to dynamic workloads in different Cloud DCs, or aggregating multiple paths provided by different service providers to achieve better performance.

Even though SDWAN has been positioned as a flexible way to reach dynamic workloads in third party data centers over multiple underlay networks, scaling becomes a major issue when there are hundreds or thousands of nodes to be interconnected by the SDWAN overlay paths.
BGP is widely used by underlay networks. This document describes using BGP to enhance the scaling properties of SDWAN overlay networks.

2. Conventions used in this document

Cloud DC: Third party data centers that host applications and workloads 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: Customer Premise Equipment

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

Homogeneous SDWAN: A type of SDWAN network in which all traffic to/from the SDWAN edge nodes has to be encrypted regardless of underlay networks. For lack of better terminology, we call this Homogeneous SDWAN throughout this document.

ISP: Internet Service Provider

NSP: Network Service Provider. NSP usually provides more advanced network services, such as MPLS VPN, private leased lines, or managed Secure WAN connections, many times within a private trusted domain, whereas an ISP usually provides plain internet services over public untrusted domains.

PE: Provider Edge

SDWAN End-point: a port (logical or physical) of a SDWAN edge node.
3. Use Case Scenario Description and Requirements

SDWAN networks can have different topologies and have different traffic patterns. To make it easier for the focused discussion in subsequent drafts on SDWAN control plane and data plane, this section describes several SDWAN scenarios that may have different need or impact to their corresponding control planes & data planes.
3.1. Requirements

3.1.1. Client Service Requirement

Client interface of SDWAN nodes can be IP or Ethernet based.

For Ethernet based client interfaces, SDWAN edge should support VLAN-based service interfaces (EVI100), VLAN bundle service interfaces (EVI200), or VLAN-Aware bundling service interfaces. EVPN service requirements are applicable to the Client traffic, as described in Section 3.1 of RFC8388.

For IP based client interfaces, L3VPN service requirements are applicable.

3.1.2. SDWAN Node Provisioning

Unlike traditional EVPN or L3VPN where PEs are deployed for long term, SDWAN edge nodes (virtual or physical) deployment at a specific location can be ephemeral. Therefore, Zero Touch Provisioning (ZTP) is a common requirement for SDWAN. ZTP for SDWAN can include many areas, but from network connectivity perspective, ZTP should include the following:

- Upon power up, an SDWAN node can reach a central SDWAN Controller (which can be burned or preconfigured in the device) via a TLS or SSL secure channel.

- The Central SDWAN Controller can designate a Local Network Controller in the proximity of the SDWAN node; the Local Network Controller and the SDWAN nodes might be connected by third party untrusted network. The Local controller does all the following 4 tasks:

  1) ZTP
  2) Auto-discovery of Network
  3) (Auto)-Provisioning for IPsec SAs (initial provisioning part)
  4) Signaling of tenant’s routes/info

BGP is well suited for (4), using Route Reflector (RR) [RFC4456] to propagate network information among SDWAN edge nodes. The SDWAN
node can establish a secure connection (TLS, SSL, etc) to the Local Network Controller (RR).

![Peer Group Diagram]

The SDWAN nodes (a.k.a. C-PEs throughout this document) belonging to the same Tenant can be far apart and can be connected by third party untrusted networks. Therefore, it is not appropriate for a SDWAN edge node (C-PE) to advertise its SDWAN Port properties to its neighbors. Each C-PE propagates its SDWAN Port attributes via the secure channel (TLS, SSL, etc.) established with the Local Controller.

C-PE-1 should include the following aspects in addition to managing client routes:
- Register the SDWAN node’s WAN port <-> local address mapping to its Local Controller. The Local Controller propagates the information to C-PE2 & C-PE3.
- Exchange IPsec property (capability such as the supported encryption algorithms, etc.) and ports NAT property (e.g. private addresses or dynamically assigned IP addresses) with the Local Controller.
- C-PE2 and C-PE3 can establish IPsec SA with the C-PE1 after receiving the information from the Local Controller.
- Then distribute the routes attached to the C-PE to its authorized peers.

Tenant separation is achieved by the Local Controller creating different Tenant based Peer Groups.
3.2. Scenarios #1: Homogeneous WAN

This is referring to a type of SDWAN network with edge nodes encrypting all traffic over WAN to other edge nodes, regardless of whether the underlay is private or public. For lack of better terminology, we call this Homogeneous SDWAN throughout this document.

Some typical scenarios for the use of a Homogeneous SDWAN network are as follows:

- A small branch office connecting to its HQ offices via the Internet. All sensitive traffic to/from this small branch office has to be encrypted, which is usually achieved using IPsec SAs.

- A store in a shopping mall may need to securely connect to its applications in one or more Cloud DCs via the Internet. A common way of achieving this is to establish IPsec SAs to the Cloud DC gateway to carry the sensitive data to/from the store.

As described in [SECURE-EVPN], the granularity of the IPsec SAs for Homogeneous SDWAN can be per site, per subnet, per tenant, or per address. Once the IPsec SA is established for a specific subnet/tenant/site, all traffic to/from the subnets/tenants/site are encrypted.
One of the key properties of homogeneous SDWAN is that the SDWAN Local Network Controller (RR) is connected to C-PEs via untrusted public network, therefore, requiring secure connection between RR and C-PEs (TLS, DTLS, etc.).

Homogeneous SDWAN has some similarity to commonly deployed IPsec VPN, albeit the IPsec VPN is usually point-to-point among a small number of endpoints and with heavy manual configuration for IPsec between end-points, whereas an SDWAN network can have a large number of end-points with an SDWAN controller to manage requiring zero touch provisioning upon powering up.

Existing Private VPNs (e.g. MPLS based) can use homogeneous SDWAN to extend over public network to remote sites to which the VPN operator does not own or lease infrastructural connectivity, as described in [SECURE-EVPN] and [SECURE-L3VPN]

3.3. Scenario #2: SDWAN WAN ports to VPN’s PEs and to Internet

In this scenario, SDWAN edge nodes (a.k.a. C-PEs) have some WAN ports connected to PEs of Private VPNs over which packets can be forwarded natively without encryption, and some WAN ports connected to the Internet over which sensitive traffic have to be encrypted (usually by IPsec SA).
In this scenario, the SDWAN edge nodes’ egress WAN ports are all IP/Ethernet based, either egress to PEs of the VPNs or to the Internet. Even if the VPN is a MPLS network, the VPN’s PEs have IP/Ethernet connections to the SDWAN edge (C-PEs). Throughout this document, this scenario is also called CPE based SDWAN over Hybrid Networks.

Even though IPsec SA can secure the packets traversing the Internet, it does not offer the premium SLA commonly offered by Private VPNs, especially over long distance. Clients need to have policies to specify criteria for flows only traversing private VPNs or traversing either as long as encrypted when over the Internet. For example, client can have those polices for the flows:

1. A policy or criteria for sending the flows over a private network without encryption (for better performance),
2. A policy or criteria for sending the flows over any networks as long as the packets of the flows are encrypted when traversing untrusted networks, or
3. A policy of not needing encryption at all.

If a flow traversing multiple segments, such as A<->B<->C<->D, has either Policy 2 or 3 above, the flow can traverse different underlays in different segments, such as over Private network underlay between A<->B without encryption, or over the public internet between B<->C in an IPsec SA.

As shown in the figure below, C-PE-1 has two different types of interfaces (A1 to Internet and A2 & A3 to VPN). The C-PEs’ loopback addresses and addresses attached to C-PEs may or may not be visible to the ISPs/NSPs. The addresses for the WAN ports can have addresses allocated by the service providers or dynamically assigned (e.g. by DHCP). One WAN port shown in the figure below (e.g. A1, A2, A3 etc.) is a logical representation of potential multiple physical ports on the C-PEs.
Some key characteristics of a Hybrid SDWAN overlay network are as follows:

- one C-PE may be connected to different ISPs/NSPs, with some of its WAN ports addresses being assigned by the ISPs/NSPs.

- The WAN ports connected to PEs of trusted private networks (e.g. MPLS VPN) hand off IP/Ethernet packets, just like today’s CPE that do not handle MPLS packets and do not participate in the underlay VPN networks’ control plane. Traffic can flow natively without encryption when be forwarded out through those WAN ports for better performance.

- The WAN ports connected to untrusted networks, e.g. the Internet, requires sensitive traffic to be encrypted, i.e. encrypted by IPsec SA.
- An SDWAN local Network Controller (RR) is connected to C-PEs via the untrusted public network, therefore, requiring secure connection between RR and C-PEs via TLS, DTLS, etc.

- The SDWAN nodes’ [loopback] addresses might not be routable nor visible in the underlay ISP/NSP networks. Routes & services attached to SDWAN edges at the SDWAN overlay layer are in different address spaces than the underlay networks.

- There could be multiple SDWAN devices sharing a common property, such as a geographic location. Some applications over SDWAN may need to traverse specific geographic locations for various reasons, such as to comply regulatory rules, to utilize specific value added services, or others.

- The underlay path selection between sites can be a local section. Some policies allow one service from CPE1 -> CPE2 -> CPE3 using one ISP/NSP underlay in the first segment (CPE1 -> CPE2), and using a different ISP/NSP in the second segment (CPE2-> CPE3).

- Services may not be congruent, i.e. the packets from A-> B may traverse one underlay network, and the packets from B -> A may traverse a different underlay.

- Different services, routes, or VLANs attached to SDWAN nodes can be aggregated over one underlay path; same service/routes/VLAN can spread over multiple SDWAN underlays at different times depending on the policies specified for the service. For example, one tenant’s packets to HQ need to be encrypted when sent over the Internet or have to be sent over private networks, while the same tenant’s packets to Facebook can be sent over the Internet without encryption.

3.4. Scenario #3: SDWAN WAN ports to MPLS VPN and the Internet

This scenario refers to existing VPN (e.g. MPLS based VPN, such as EVPN or IPVPN) adding extra ports facing untrusted public networks allowing PEs to offload some low priority traffic to those ports facing public networks when the VPN MPLS paths are congested. Throughout this document, this scenario is also called Internet Offload for Private VPN, or PE based SDWAN.
In this scenario, it is important that the packets offloaded to untrusted public network be encrypted. In this scenario, there is a secure BGP connection between RR & PEs.

PE based SDWAN can be used by VPN service providers to temporarily increase bandwidth between sites when they are not sure if the demand will sustain for long period of time or as a temporary solution before the permanent infrastructure is built or leased.

Here are some key properties for PE based SDWAN:

- For MPLS based VPN, PEs continue having MPLS encapsulation handoff to existing paths.
- The BGP RR is connected to PEs in the same way as VPN, i.e. via the trusted network.
- For the added Internet ports, PEs have IP packets handoff, i.e. sending and receiving IP data frames. Internally, PEs can have the option to encapsulate the MPLS payload in IP, as specified by [RFC4023].
- The ports facing public internet might get IP addresses assigned by ISPs, which may not be in the same address domain as PEs.'
- Ports facing public internet are not as secure as the ports facing private infrastructure. There could be spoofing, or DDOS attacks to the ports facing public internet. Extra consideration must be given when injecting the new routes from public network into VRFs.
- Even though packets are encrypted over public internet, the performance SLA is not guaranteed over public internet. Therefore, clients may have policies only allowing some flows to be offloaded to internet path.

4. Provisioning Model

4.1. Client Service Provisioning Model

The provisioning tasks described in Section 4 of RFC8388 are the same for the SDWAN client traffic. When client traffic are multi-homed to two (or more) C-PEs, the Non-Service-Specific parameters need to be provisioned per the Section 4.1.1 of RFC8388.

Since most SDWAN nodes are ephemeral and have small number of IP subnets or VLANs attached to the client ports of the SDWAN nodes, it is recommended to have default and simplified Service-specific parameters for each client port, remotely managed by the SDWAN Network Controller (i.e. the RR) via the secure channel (TLS/DTLS) between the controller and the C-PEs.

More details are to be added.

4.2. WAN Ports Provisioning Model

Since the deployment of PEs to MPLS VPN are for relatively long term, the common provisioning procedure for PE’s WAN ports is via CLI.

A SDWAN node deployment can be ephemeral and its location can be in remote locations, manual provisioning for its WAN ports is not acceptable. In addition, a SDWAN WAN port’s IP address can be dynamically assigned or using private addresses. Therefore, it is necessary to have a separate control protocol; something like NHRP did for ATM, for a SDWAN node to register its WAN property to its controller dynamically.

Unlike a PE to MPLS based VPN where its WAN ports are homogeneously facing MPLS private network and all traffic are egressed in MPLS data frames through its WAN ports, the WAN ports of a SDWAN node can be connected to a PE of VPN, MPLS private network directly, the public Internet, or the various combinations of all.

For Scenario #1 above, the WAN ports can face public internet or VPN.

For Scenario #2 above, WAN ports are either configured as connecting to PEs of VPN where traffic can be sent as IP/Ethernet without encryption, or configured as connecting to public Internet.

For Scenario #3 above, the WAN ports are either configured as VPN egress ports (hand off MPLS data frames), or as connecting to the public internet that requires MPLS in IP in IPsec encapsulation.

4.2.1. Why BGP as Control Plane for SDWAN WAN Ports Registration?

For a small sized SDWAN network, traditional hub & spoke model using NHRP or DSVPN/DMVPN with a hub node (or controller) managing SDWAN node WAN ports mapping (e.g. local & public addresses and tunnel identifiers mapping) can work reasonably well. However, for a large SDWAN network, say more than 100 nodes with different types of topologies, the traditional approach becomes very messy, complex and error prone.

Here are some of the compelling reasons of using BGP instead of extending NHRP/DSVPN/DMVPN. (Same as the reasons quoted by LSVR on why using BGP):

- BGP already widely deployed as sole protocol (see RFC 7938)
- Robust and simple implementation
- Wide acceptance - minimal learning
- Reliable transport
- Guaranteed in-order delivery
- Incremental updates
- Incremental updates upon session restart
- No flooding and selective filtering
- RR already has the capability to apply policies to communications among peers.

5. SDWAN Traffic Forwarding Walk Through

BGP based EVPN control plane are still applicable to routes attached to the client ports of SDWAN nodes. Section 5 of RFC8388 describes the BGP EVPN NLRI Usage for various routes of client traffic. The procedures described in the Section 6 of RFC8388 are same for the SDWAN client traffic.

The only additional consideration for SDWAN is to control how traffic egress the SDWAN edge node to various WAN ports.

5.1. SDWAN Network Startup Procedures

A SDWAN network can add or delete SDWAN edge nodes on regular basis depending on user requests.

- For Scenario #1: a SDWAN edge node in a shopping mall or Cloud DC can be added or removed on demand. The Zero Touch Provisioning described in 3.1.2 are required for the node startup.
- For Scenario #2: this can be Data Centers or enterprises upgrading their CPEs to add extra bandwidth via public internet in addition to VPN services that they already purchased. Before the node powers up or upgraded, there should be links connected to the PEs of a provider VPNs.
- For Scenario #3, the Internet facing WAN ports are added to (or removed from) existing VPN PEs.

5.2. Packet Walk-Through for Scenario #1

Upon power up, a SDWAN node can learn client routes from the Client facing ports, in the same way as EVPN described in RFC8388. Controller facilitates the IPsec SA establishment and rekey management as described in [SECURE-EVPN]. Controller manages how client’s routes are associated with individual IPsec SA.

[SECURE-L3VPN] describes how to extend the RFC4364 VPN to allow some PEs being connected to other PEs via public networks. [SECURE-L3VPN] introduces the concept of Red Interface & Black Interface on those
PEs, with RED interfaces face clients’ routes within the VPN and the Black Interfaces being WAN ports over which only IPsec-protected packets to the Internet or other backbone network are sent so that eliminating the need for MPLS transport in the backbone.

[SECURE-L3VPN] assumes PEs terminate MPLS packets, and use MPLS over IPsec when sending over the Black Interfaces.

[SECURE-EVPN] describes a solution where BGP point-to-multipoint signaling is leveraged as control plane for SDWAN Scenario #1. It utilizes the BGP RR to facilitate the key and policy exchange among PE devices to create private pair-wise IPsec Security Associations without IKEv2 point-to-point signaling or any other direct peer-to-peer session establishment messages.

When C-PEs do not support MPLS, the approaches described by RFC8365 can be used, with addition of IPsec encrypting the IP packets when sending packets over the Black Interfaces.

5.3. Packet Walk-Through for Scenario #2

In this scenario, C-PEs have some WAN ports connected to the public internet and some WAN ports connected to (i.e. directly connected to) PEs of trusted VPN. The C-PEs in Scenario #2 are almost like CPEs to MPLS VPN that have the IP/Ethernet data frames egress to the PEs of the VPN, except the packets need encryption if egress to the WAN ports facing public Internet.

Users specify the policy or criteria on which flows can only egress WAN ports facing trusted VPN without encryption, which can egress the WAN ports facing the public Internet with encryption, or which can egress WAN ports facing the public Internet without encryption.

The Control Plane should not learn routes from the Public Network facing WAN ports. Should strictly follow the policies specified by the users. The internet facing WAN ports can face potential DDoS attacks, additional anti-DDoS mechanism has to be implemented on WAN ports facing those public networks.

The Scenario #2 SDWAN Control Plane has three distinct functional components:
- SDWAN node’s WAN ports property registration to the SDWAN Network Controller (BGP RR).
  - This is used to inform the SDWAN controller of all the underlay networks to which the C-PE is connected.
  - RR is responsible for propagating the C-PE WAN ports properties to authorized peers.

- Controller Facilitated IPsec SA management and NAT information distribution
  - Used by the SDWAN controller to facilitate or manage the IPsec configurations and peer authentications for all IPsec SAs terminated at the SDWAN nodes.
  - When WAN ports have private addresses, need exchange between SDWAN edges and the RR about the type of NAT, and mapping of the private addresses/ports <-> public addresses/ports.

- Attached routes distribution via BGP RR, which can be EVPN, IPVPN or others.
  - This is for the overlay layer’s route distribution, so that a C-PE can establish the overlay routing table that identifies the next hop for reaching a specific route/service attached to remote nodes. [SECURE-EVPN] describes EVPN and other options.
5.3.1. SDWAN node WAN Ports Properties Registration

In Figure 6, A1/A2/A3/B1/B2/B3 WAN ports can be from different network providers.

```
+----+  +---------+  packets encrypted over  +---------+  +----+
| CN3|--|         A1-----+ Untrusted    +------ B1       |--| CN1|
+----+  | C-PE1   A2-----+              +-------B2 C-PE2 |  +----+
    | 10.1.1.1 |                             |10.1.2.1|
    | 10.1.1.1 |                             |10.1.2.1|
+----+  |         |   +--+              +---+   |        |  +----+
| CN2|--|         A3  |PE+--------------+PE |---B3       |--| CN3|
+----+  +---------+   +--+   trusted    +---+   +--------+  +----+
    |     VPN      |
    +--------------+
```

Figure 6: SDWAN Scenario #2 WAN Ports Registration

Each SDWAN edge (C-PE) needs to register its WAN ports properties along with its Loopback addresses to the SDWAN Network Controller (RR). The policies that govern the communications among peers are managed and controlled by the SDWAN Controller. Individual SDWAN edge relies on its SDWAN Controller to determine which peers can establish connections. The SDWAN controller is responsible for propagating the mapping information to the authorized peers. If C-PE-1 is not authorized to communicate with C-PE-n, C-PE-1’s WAN port<->Loopback address mapping will not be propagated to C-PE-n.

A C-PE’s Loopback addresses & attached routes may not be visible to some ISPs/NSPs to which the CPE’s WAN port is connected.

5.3.2. Controller Facilitated IPsec SA & NAT management

One IPsec SA between two end points is straightforward. However, for a network with many IPsec SAs among many end points, the configuration and IPsec Key management for the entire network can be complex.
For a 1,000-node network, each node is responsible for maintaining and managing 999 keys to all their peers, which could potentially result in 1,000,000 key exchanges to authenticate among all nodes. In addition, when an edge node has multiple tenants attached, the edge node may need to establish multiple tunnels for tenants. For example, for a network with N nodes, a node A has 5 tenants attached to it, then the node A has to maintain 5*(N-1) number of keys if each tenant needs to communicate with all other nodes.

In addition, all the IPsec keys have to be refreshed periodically, which adds more complexity. Therefore, simplification facilitated by an SDWAN controller is necessary for large-scale SDWAN deployment.

When the SDWAN IPsec SAs are fine-grained, such as per client address, per client’s VLAN, the number of IPsec SAs & Keys to be managed can go much higher, leading to more IPsec management complexity. It is better to aggregate multiple flows into one IPsec SA.

SDWAN edge nodes can rely on the SDWAN controller to facilitate the pair-wise IPsec key establishment and refreshment [RFC7296] and maintain the Security Policy Database (SPD) [RFC4301].

- In the Figure 5 SDWAN Scenario #2 above, if C-PE1 & C-PE2 each has two ports facing two different ISPs networks, and their loopback addresses are not visible to the ISPs, i.e. the C-PE1 & C-PE2 are using a provider assigned IP addresses for A1/A2/B1/B2; you are going to need minimum four IPsec SAs between C-PE1 & C-PE2.
- When C-PEs loopback addresses are visible to ISPs/NSPs, i.e. the C-PEs’ private source and destination IPs are part of a prefix exported to the ISP(s) in each site, it is possible to have one IPsec SA between C-PE1 & C-PE2.

The IP addresses of SDWAN WAN port can be dynamic (e.g. assigned by DHCP) or private IP. Some SDWAN nodes are identified by "System-ID" or Loopback addresses that are only locally significant. In some SDWAN environments, "System-ID + PortID" are used to uniquely identify a SDWAN WAN port. Sometimes, a SDWAN tunnel end-point can be associated with "private IP" + "public IP" (if NAT is used.)

When CPE WAN ports are private addresses, an additional sub-TLV has to be added to the [Tunnel-Encap] to describe the additional
information about the NAT property of SDWAN nodes’ WAN ports. A
SDWAN 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 the authorized peers via the SDWAN Controller.

5.3.3. BGP Based SDWAN client routes

The client routes attached to SDWAN client ports have to be
distributed to all SDWAN edge nodes, just like BGP/MPLS IP VPN
[RFC4364], so that all SDWAN edges can establish the overlay routing
table that identifies the remote SDWAN edges to reach a specific
route/service. When C-PEs do not handle MPLS, RFC8365 can be used
for packets over WAN ports, albeit applying IPsec SA encryption when
sent over the WAN ports facing the public networks.

Using the terminologies described by [SECURE-L3VPN], the RED
interface are the clients’ ports and the ports facing private
networks (e.g. connected to the PEs of MPLS VPN). Black Interfaces
are ports facing public networks. The behavior described in [SECURE-
L3VPN] applies to this scenario too, the C-PEs cannot mix the routes
learned from the Black Interfaces with the Routes from RED
Interfaces.

To minimize the burden on SDWAN edge nodes (especially low powered
virtual SDWAN edges), some SDWAN network can let SDWAN controller
take care of authenticating communications among SDWAN edge nodes
instead of pushing down policies to SDWAN edge nodes. SDWAN Edge
nodes might get clients routes from SDWAN controller instead of
learning from clients ports.

The Hybrid SDWAN control plane for distributing clients’ routes is
more similar to overlay using EVPN [RFC8365], albeit the packets
sent over the internet facing ports have to be encrypted by IPsec
SA.

[Tunnel-Encap] can be used to associate client routes with specific
tunnels:

- C-PE1 can advertise the following properties to others C-PEs
  via RR:
  - Encapsulation capability of the Ports to VPN PE
  - Encapsulation capability of the Ports to the Internet:
    GRE-IPsec, or MPLS over GRE over IPsec
- with prior established IPsec SA
- NAT information if ports are private addresses

- The Remote Endpoint sub-TLV is NOT appropriate because
  - The network to which a SDWAN port is connected might have an identifier that is more than the AS number. The SDWAN controller might use its own specific identifier for the network.
  - Suggest using an SDWAN overlay specific Transport-Network-ID to represents the connected networks.

The underlay network selections to next hop C-PE can be a local decision. Different services, routes, or VLANs can be aggregated to one underlay network between two C-PEs; the same service/routes/VLAN can spread over multiple SDWAN underlay networks at the next segment.

5.4. Packet Walk-Through for Scenario #3

The behavior described in [SECURE-L3VPN] applies to this scenario, except C-PEs not only have RED interfaces facing clients within the VPN but also have RED interface facing MPLS backbone, with additional BLACK interfaces facing the untrusted public networks. The C-PEs cannot mix the routes learned from the Black Interfaces with the Routes from RED Interfaces. The routes learned from core-facing RED interfaces are for underlay and cannot be mixed with the routes learned over access-facing RED interfaces that are for overlay. Furthermore, the routes learned over core-facing interfaces (both RED and BLACK) can be shared in the same GLOBAL route table.

There may be some added risks of the packets from the ports facing the Internet. Therefore, special consideration has to be given to the routes from WAN ports facing the Internet. RFC4364 describes using an RD to create different routes for reaching same system. A similar approach can be considered to force packets received from the Internet facing ports to go through special security functions before being sent over to the VPN backbone WAN ports.
6. Manageability Considerations

SDWAN overlay networks utilize the SDWAN controller to facilitate route distribution, central configurations, and others. To minimize the burden on SDWAN edge nodes, SDWAN Edge nodes might not need to learn the routes from clients.

7. Security Considerations

Having WAN ports facing the public Internet introduces the following security risks:

1) Potential DDoS attack to the C-PEs with ports facing internet.

2) Potential risk of provider VPN network being injected with illegal traffic coming from the public Internet WAN ports on the C-PEs.

8. IANA Considerations

None

9. References

9.1. Normative References


9.2. Informative References


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


10. Acknowledgments

Acknowledgements to Jim Guichard, John Scudder, Darren Dukes, Andy Malis and Donald Eastlake for their review and contributions.

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

Linda Dunbar
Huawei
Email: ldunbar@futurewei.com

James Guichard
Huawei
Email: james.n.guichard@futurewei.com

Ali Sajassi
Cisco
Email: sajassi@cisco.com

John Drake
Juniper
Email: jdrake@juniper.net