Network Working Group
Internet Draft
Intended status: Informational
Expires: October 29, 2020

L. Dunbar
J. Guichard
Futurewei
Ali Sajassi
Cisco
J. Drake
Juniper
B. Najem
Bell Canada
Ayan Barnerjee
D. Carrel
Cisco

April 29, 2020

BGP Usage for SDWAN Overlay Networks
draft-dunbar-bess-bgp-sdwan-usage-07

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 demonstrate how BGP-based control plane is used for large scale SDWAN overlay networks with little manual intervention.

SDWAN edge nodes are commonly interconnected by multiple underlay networks which can be owned and managed by different network providers.

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 October 29, 2020.

Table of Contents

1. Introduction

There are three key characteristics of "SDWAN" networks:

- Augment of transport, which refers to utilizing overlay paths over different underlay networks. Very often there are multiple parallel overlay paths between any two SDWAN edges, some of which are private networks over which traffic can traverse with or without encryption, others require encryption, e.g. over untrusted public networks.
- Enable direct Internet access from remote sites, instead hauling all traffic to Corporate HQ for centralized policy control.
- Some traffic are routed based on application IDs instead of based on destination IP addresses.

[Net2Cloud-Problem] describes the network related problems that enterprises face to connect enterprises' branch offices to dynamic workloads in different Cloud DCs, including using SDWAN to aggregate multiple paths provided by different service providers to achieve

better performance and to accomplish application ID based forwarding.

Even though SDWAN has been positioned as a flexible way to reach dynamic workloads in third party Cloud data centers over different underlay networks, scaling becomes a major issue when there are hundreds or thousands of nodes to be interconnected by an SDWAN overlay networks.

BGP is widely used by underlay networks. This document describes using BGP for edge nodes to exchange information across the 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.

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, traffic can traverse without additional encryption; when the underlay networks are public, such as the Internet, some traffic may need to be encrypted when traversing through (depending on user provided policies).

SDWAN IPsec SA: IPsec Security Association between two SDWAN ports or nodes.

SDWAN over Hybrid Networks: SDWAN over Hybrid Networks typically have edge nodes utilizing bandwidth resources from multiple service providers. In Hybrid SDWAN network, packets over private networks can go natively without encryption and are encrypted over the untrusted network, such as the public Internet.

WAN Port: A Port or Interface facing an ISP or Network Service Provider (NSP), with address (usually public routable address) allocated by the ISP or the NSP.

C-PE: SDWAN Edge node, which can be CPE for customer managed SDWAN, or PE that is for provider managed SDWAN services).

ZTP: Zero Touch Provisioning

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 impact on their corresponding control planes & data planes.

3.1. Requirements

3.1.1. Supporting Multiple SDWAN Segmentations

The term "network segmentation", a.k.a. SDWAN instances, is referring to the process of dividing the network into logical sub-networks using isolation techniques on a forwarding device such as a switch, router, or firewall. For a homogeneous network, such as MPLS VPN or Layer 2 network, VRF or VLAN are used to achieve the network segmentation.

As SDWAN is an overlay network arching over multiple types of networks, VRF or VLAN can't be used directly to differentiate SDWAN network segmentations.

However, BGP already has the capability to differentiate SDWAN segmentations:

- Create a SDWAN Target ID in the BGP Extended Community to represent different SDWAN Segmentations
  - Same as Route Target, just use a different name to differentiate from VPN if a CPE supports traditional VPN with multiple VRFs and supports multiple SDWAN Segmentations (instances).
- When the SDWAN Target ID is used,
  - Use the similar approach as VPN Label carried by NLRI Path Attribute [RFC8277] to identify routes belonging to different SDWAN Segmentations.
  - The MPLS VPN SAFI 128 & Route Distinguisher can be used for routes belonging to different SDWAN instances.

3.1.2. 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 the Section 3.1 of RFC8388.

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

3.1.3. Application Flow Based Segmentation

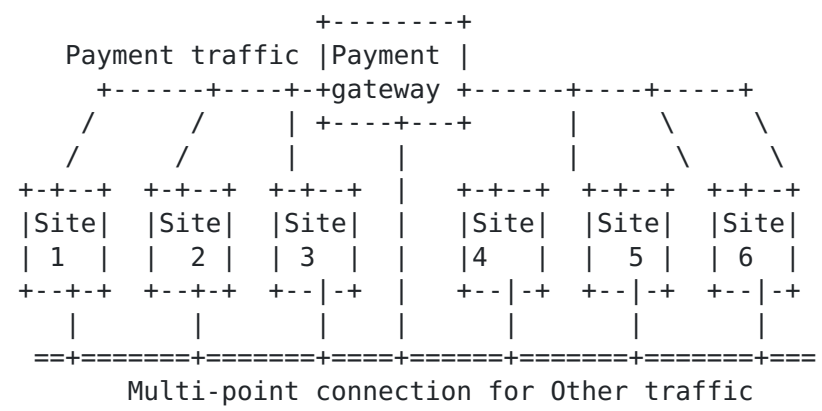
Application Flow based Segmentation, also known as SDWAN Traffic Segmentation, enables the separation of the traffic based on the business and the security needs for different users' groups and/or application requirements. Each user group and/or applications may need different isolated topology and/or policies to fulfill the business requirements.

The Application Flow based Segmentation concept is analogous to VLAN (in L2 network) and VRF (in L3 network).

One can think about the Application Flow based Segmentation as a feature that can be provided or enabled on a single SDWAN service (or domain) to a single Subscriber. Each SDWAN Service can have one or more overlay Segments to support the business requirement; each Segment has its own policy, topology and application/user groups. Applications/users' group can belong to more than one Segment.

For example, a retail business requires the point-of-sales (PoS) application in all stores to be isolated from other applications AND routed only to the payment processing entity at a hub site (i.e. hub and spoke); however, the same retail business requires the other applications to be routed to all sites (i.e. multipoint-to multipoint) AND isolated from the PoS application.

In the figure below, the traffic from the PoS application follows a Tree topology, whereas other traffic can be multipoint-to-multipoint topology.



Another example is an enterprise who wants to isolate the traffic for each department and have different topology and policy for different department; the HR department may need to access certain applications that are NOT accessible by the engineering department. In addition, the contractors may have a limited access to the enterprise resources.

3.1.4. Zero Touch Provisioning

Unlike traditional EVPN or L3VPN whose 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), or Plug and Play, is a common requirement for SDWAN. When an SDWAN edge is physically installed at a location or instantiated on a VM in a Cloud DC, ZTP automates follow-up steps, including updates to the OS, software version, and configuration prior to connection. From network control perspective, ZTP includes the following:

- Upon power up, an SDWAN node can establish transport layer secure connection (such as TLS, SSL, etc.) to its controller whose address can be burned or preconfigured on the device.

- The SDWAN Controller can designate a Local Network Controller in the proximity of the SDWAN node; the Local Network Controller manages and monitor the communication policies of the edge node.

3.1.5. Constrained Propagation of SDWAN Edge Properties

One SDWAN edge node may only be authorized to communicate with a small number of other SDWAN edge nodes. Under this circumstance, the property of the SDWAN edge node cannot be propagated to any other nodes who are not authorized to communicate. But a remote SDWAN edge node upon powering up might not have the proper policies to know who the authorized peers are. Therefore, it is very essential for SDWAN deployment have a central point to distribute the properties of each SDWAN edge node to its authorized peers.

BGP is well suited for this purpose. RFC 4684 has specified the procedure to constrain the distribution of BGP UPDATE to only a subset of SDWAN edges. Basically, each edge node informs the Route Reflector (RR) [RFC4456] on its interested SDWAN instances. The RR only propagates the BGP UPDATE for the relevant SDWAN instances to the edge.

Usually the connection between a SDWAN edge node and its RR is over insecure network. Therefore, upon power up, a SDWAN node needs to establish a secure transport layer connection (TLS, SSL, etc.) to its designated RR. The BGP UPDATE messages need to be sent over the secure channel (TLS, SSL, etc.) to the RR.

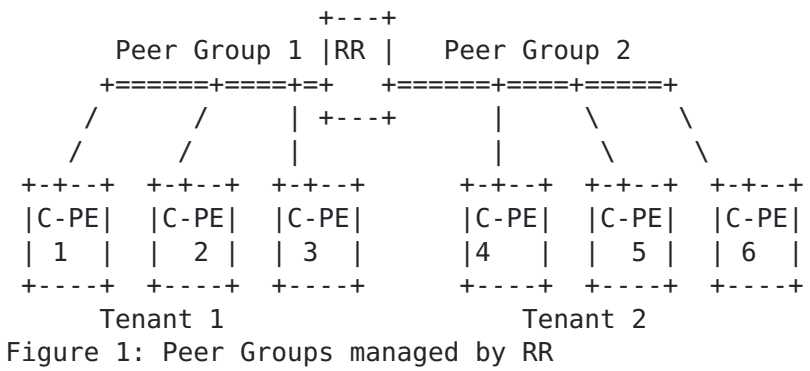


Figure 1: Peer Groups managed by RR

Tenant separation is achieved by the RR creating different Tenant based Peer Groups.

3.2. Scenarios #1: Homogeneous WAN

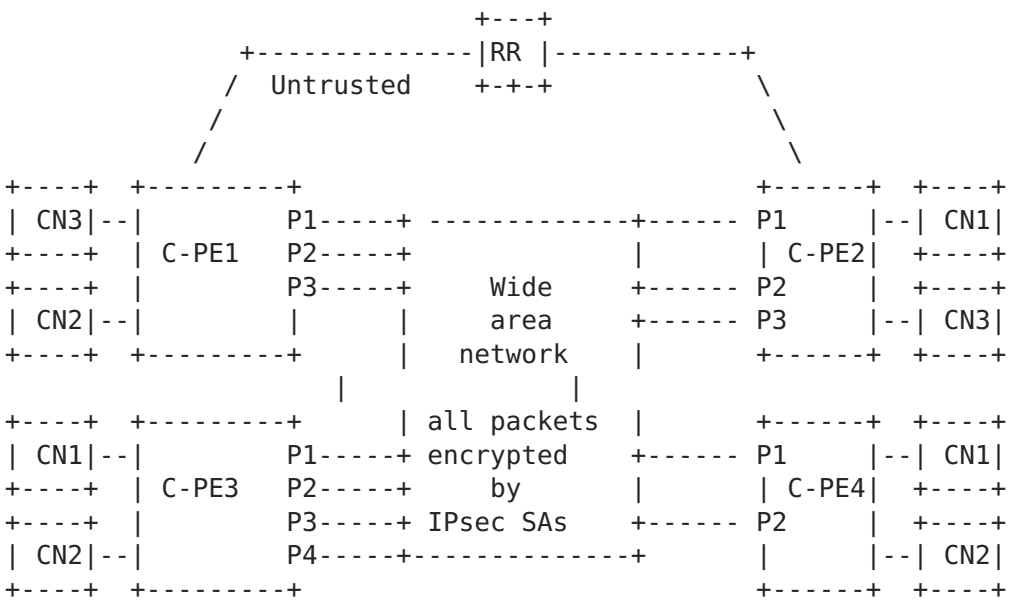
This is referring to a type of SDWAN network with edge nodes encrypting all traffic over WAN to other edge nodes, regardless of whether the underlay is private or public. For lack of better terminology, we call this Homogeneous SDWAN throughout this document.

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.



CN: Client Networks, which is same as Tenant Networks used by NVo3

Figure 2: Homogeneous SDWAN

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).

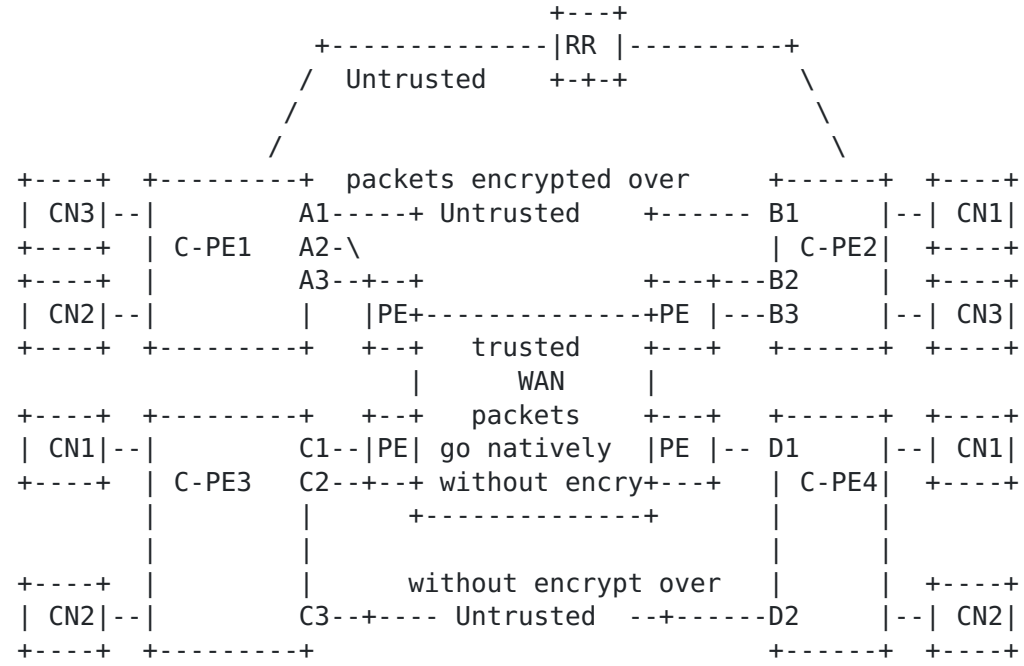
In this scenario, the SDWAN edge nodes' egress WAN ports are all IP/Ethernet based, either egress to PEs of the VPNs or to the Internet. Even if the VPN is a MPLS network, the VPN's PEs have IP/Ethernet connections to the SDWAN edge (C-PEs). Throughout this document, this scenario is also called CPE 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 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.



CN: Client Network

Figure 3: Hybrid SDWAN

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 different 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 with regulatory rules, to utilize specific value added services, or others.

- The underlay path selection between sites can be a local decision. Some policies allow one service from C-PE1 -> C-PE2 -> C-PE3 using one ISP/NSP underlay in the first segment (C-PE1 -> C-PE2) and using a different ISP/NSP in the second segment (C-PE2-> 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 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, the packets offloaded to untrusted public network must be encrypted.

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.

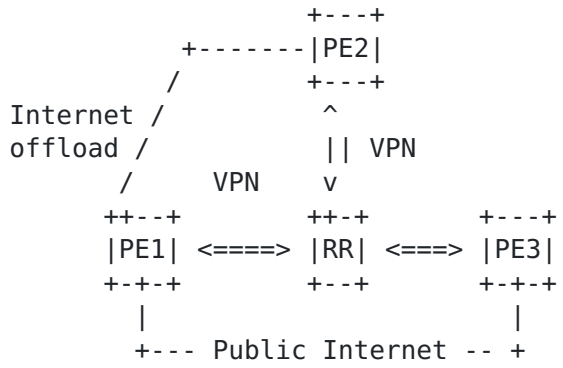


Figure 4: Additional Internet paths added to the VPN

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 learned 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. BGP Walk Through

4.1. BGP Walk Through for Homogeneous SDWAN

In the figure below, packets destined towards multiple routes attached to the C-PE2 can be carried by one IPsec tunnel. Then one BGP UPDATE can be announced by C-PE2 to its RR.

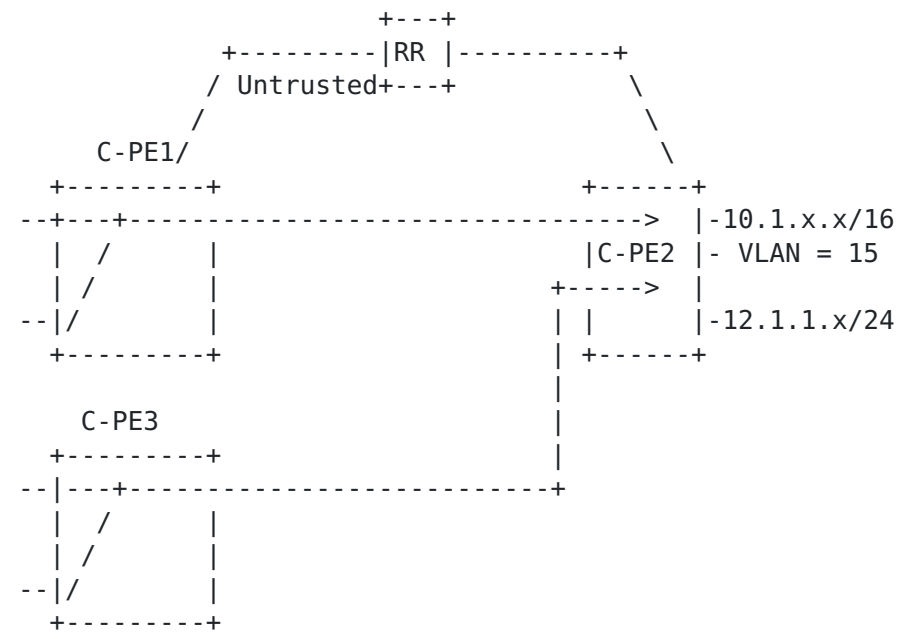


Figure 5: (see *.pdf for more accurate figure)

The BGP UPDATE Message from C-PE2 to RR should have the client routes encoded in the MP-NLRI Path Attribute and the IPsec Tunnel associated information encoded in the Tunnel-Encap Path Attributes as described in the [SECURE-EVPN]:

- MP-NLRI Path Attribute: to indicate multiple routes attached to the C-PE2:
  10.1.x.x/16
  VLAN #15
  12.1.1.x/24
- Tunnel-Encap Path Attribute: to describe the IPsec attributes for routes encoded in the NLRI Path Attribute:
  IPsec attributes for remote nodes to establish the IPsec tunnel to C-PE2.

If different client routes attached to C-PE2 needs to be reached by separate IPsec tunnels, then multiple BGP UPDATE messages need to be sent to the remote nodes. If C-PE2 doesn't have the policy on mapping of clients' routes to IPsec tunnels, RR needs to check the client routes policies to send separate BGP UPDATE messages to the remote edge nodes.

There could be policies governing the topologies of a client's different routes attached to an edge node. For example, VLAN #25 and

route 22.1.1.x/24 could be the Payment Applications described in the Section 3.1.2 that can only communicate with Payment Gateway attached to C-PE3. If C-PEs don't have the policy to govern the communication peers, RR can take over the responsibility of only send BGP UPDATE for the Blue routes to C-PE1 and send the RED routes to C-PE3.

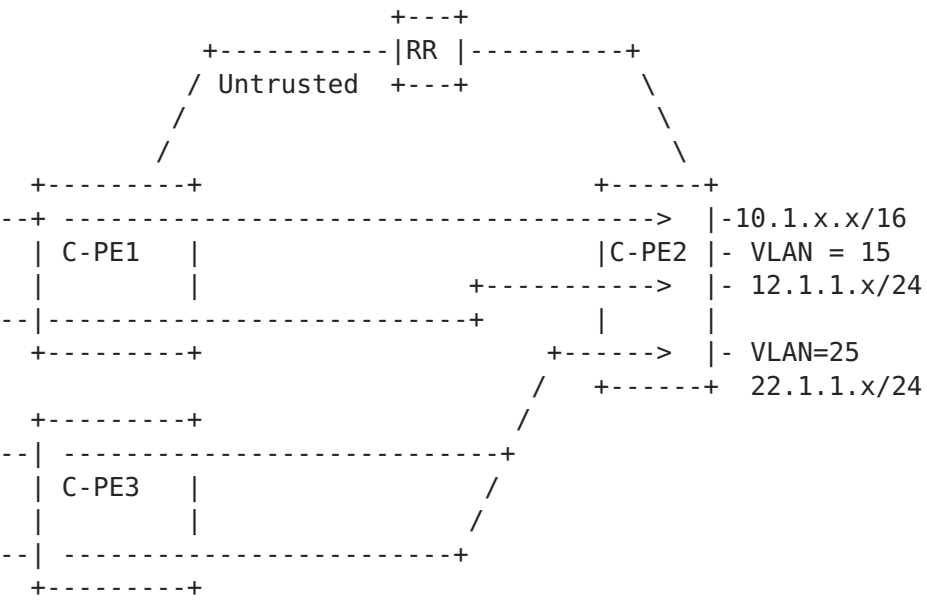


Figure 6: (see *.pdf for more accurate figure)

UPDATE 1:
- MP-NLRI Path Attribute:
  10.1.x.x/16
  VLAN #15
  12.1.1.x/24

- Tunnel-Encap Path Attribute:
  IPsec SA attributes for IPsec tunnels to C-PE2 from any node for reaching 10.1.x.x/16, VLAN #15, and 12.1.1.x/24.

UPDATE 2 (only sent to C-PE3)
- MP-NLRI Path Attribute:
  VLAN #25
  22.1.1.x/24

- Tunnel-Encap:
  IPsec SA attributes for IPsec tunnels to C-PE2 from C-PE3 for reaching VLAN #25 and subnet 22.1.1./24.

4.2. BGP Walk Through for Application Flow Based Segmentation

If the applications are assigned with unique IP addresses, the Application Flow based Segmentation described in Section 3.1.2 can be achieved by advertising different BGP UPDATE messages to different nodes. In the Figure below, the following BGP Updates can be advertised to ensure that Payment Application only communicates with the Payment Gateway:

BGP UPDATE #1 from C-PE2 to RR for the RED P2P topology (only propagated to Payment GW node:

- MP-NLRI Path Attribute:
  - 30.1.1.x/24
- Tunnel Encap Path Attribute
  - IPsec Attributes for PaymentGW ->C-PE2

BGP UPDATE #2 from C-PE2 to RR for the routes to be reached by Purple:

- MP-NLRI Path Attribute:
  - 10.1.x.x
  - 12.4.x.x
- TunnelEncap Path Attribute:
  - Any node to C-PE2

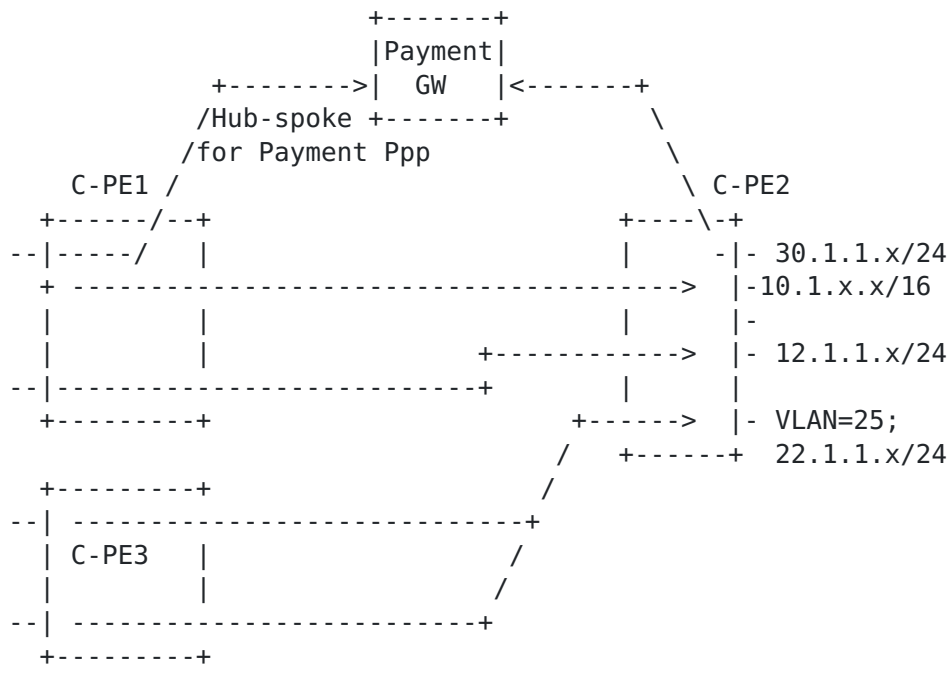


Figure 7: (see *.pdf for more accurate figure)

4.3. Client Service Provisioning Model

The provisioning tasks described in Section 4 of RFC8388 are the same for the SDWAN client traffic. When client traffic is 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 their client ports, it is recommended to have default and simplified Service-specific parameters for each client port, remotely managed by the SDWAN Network Controller via the secure channel (TLS/DTLS) between the controller and the C-PEs.

4.4. 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 with Ethernet/IP, MPLS private network directly via MPLS headers, or the public Internet.

For Scenario #1 described in Section 3.2, the WAN ports can face public internet or VPN.

For Scenario #2 described in Section 3.3, 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 that requires encryption for packets egress out.

For Scenario #3 described in Section 3.4, 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.

4.5. Why BGP as Control Plane for SDWAN?

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 has the built-in capability to constrain the propagation of SDWAN edge node properties to a small number of edge nodes [RFC4684].

- RR already has the capability to apply policies to communications among peers.

- BGP is 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

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. RED interfaces face the VPN over which packets can be forwarded natively without encryption. Black Interfaces face public network over which only IPsec-protected packets are forwarded.

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

[SECURE-EVPN] describes a solution 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 with direct connect to PEs of trusted VPN. The C-PEs in Scenario #2 have the plain IP/Ethernet data frames egress to the PEs of the VPN, encrypted data frames egress the WAN ports facing the public Internet.

Users specify the policy or criteria on which flows can only egress WAN ports facing the 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 internet facing WAN ports can face potential DDoS attacks, additional anti-DDoS mechanism has to be enabled on those WAN ports and the Control Plane should not learn routes from the Public Network facing WAN ports.

The Scenario #2 SDWAN Control Plane should include those three distinct functional components:

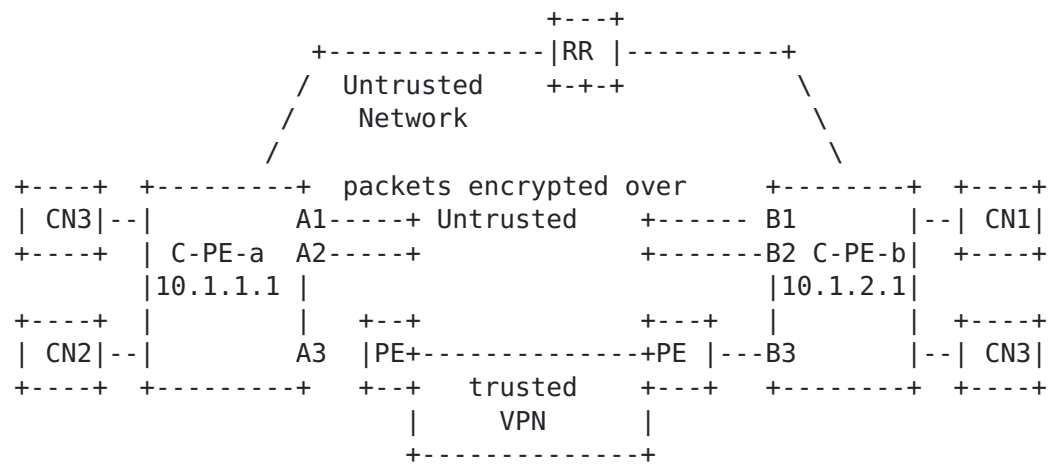


Figure 8: SDWAN Scenario #2

- SDWAN node's WAN ports property registration to the SDWAN Network Controller.
  - o See 5.3.1. for detail.

- Controller Facilitated IPsec SA management and NAT information distribution
  - o See 5.3.2. for details.

- Attached routes distribution via BGP, which can be EVPN, IPVPN or others.
  - o This is for the clients' 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.

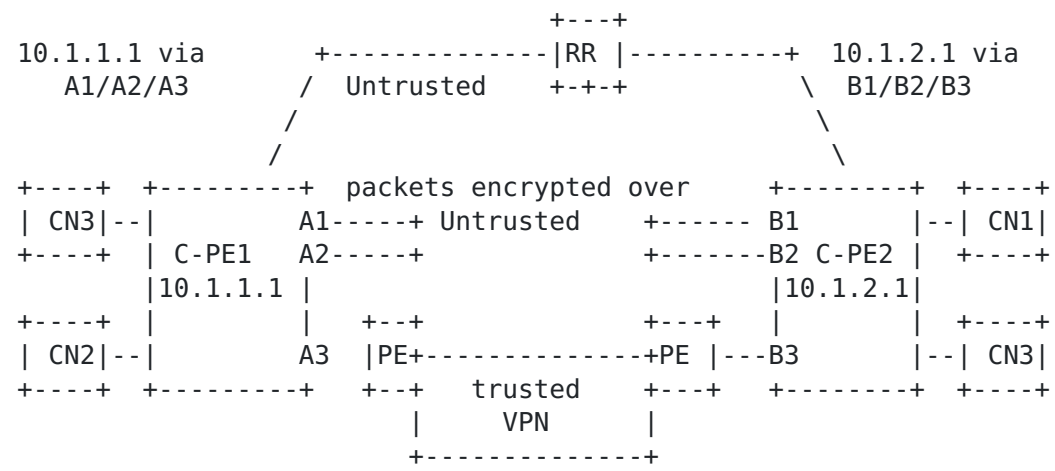


Figure 9: SDWAN Scenario #2 WAN Ports Registration

Each SDWAN edge(C-PE) needs to register its WAN ports properties along with its Loopback addresses to the SDWAN Network Controller. 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.

Section 4 describes how C-PEs use the BGP UPDATE messages to propagate their local information to their corresponding RR.

5.3.2. Controller Facilitated IPsec SA & NAT management

Setting up and managing one IPsec SA between two points is straightforward and simple. But managing multi-point IPsec SAs among many points can be overwhelming. For a 1,000-node network, each node

may need to manage 999 IPSec SA 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 to isolate traffic from different tenants. 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.

All the IPsec keys have to be refreshed periodically. Therefore, simplification facilitated by an SDWAN controller is necessary for large-scale SDWAN deployment.

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].

- For the SDWAN Scenario #2 in described in Section 3.3. , if C-PE1 & C-PE2 loopback addresses are not visible to the public network's ISP(s). C-PE1 & C-PE2 can use their provider assigned IP addresses for WAN ports A1/A2/B1/B2 to establish WAN Port based IPsec SA through the public network. Under this circumstance, there need to be minimum four IPsec SAs between C-PE1 & C-PE2 internet facing WAN ports.
- 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.

5.4. 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 but also have RED interface facing MPLS backbone, with additional BLACK interfaces facing the untrusted public networks for the WAN side. 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

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

[RFC4364] E. rosen, Y. Rekhter, "BGP/MPLS IP Virtual Private networks (VPNs)", Feb 2006.

[RFC7296] C. Kaufman, et al, "Internet Key Exchange Protocol Version 2 (IKEv2)", Oct 2014.

[RFC7432] A. Sajassi, et al, "BGP MPLS-Based Ethernet VPN", Feb 2015.

[RFC8365] A. Sajassi, et al, "A network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", March 2018.

9.2. 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.

[BGP-SDWAN-Port] L. Dunbar, H. Wang, W. Hao, "BGP Extension for SDWAN Overlay Networks", draft-dunbar-idr-bgp-sdwan-overlay-ext-03, work-in-progress, Nov 2018.

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

[SECURE-EVPN] A. Sajassi, et al, "Secure EVPN", draft-sajassi-bess-secure-evpn-01, Work-in-progress, March 2019.

[SECURE-L3VPN] E. Rosen, R. Bonica, "Secure Layer L3VPN over Public Infrastructure", draft-rosen-bess-secure-l3vpn-00, Work-in-progress, June 2018.

[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.

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
Futurewei
Email: ldunbar@futurewei.com

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

Ali Sajassi
Cisco
Email: sajassi@cisco.com

John Drake
Juniper
Email: jdrake@juniper.net

Basil Najem
Bell Canada
Email: basil.najem@bell.ca

David Carrel
Cisco
Email: carrel@cisco.com

Ayan Banerjee
Cisco
Email: ayabaner@cisco.com