Network Working Group
Internet Draft
Intended status: Informational
Expires: June 28, 2025

L. Dunbar
Futurewei
A. Sajassi
Cisco
J. Drake
Independent
B. Najem
Bell Canada
S. Hares

July 28, 2025

BGP Usage for SD-WAN Overlay Networks
draft-ietf-bess-bgp-sdwan-usage-26

Abstract

This document explores the complexities involved in managing large scale Software Defined WAN (SD-WAN) overlay networks, along with various SD-WAN scenarios. Its objective is to illustrate how a BGP-based control plane can effectively manage these overlay networks by distributing edge service reachability information, WAN port attributes, and underlay path details, thereby minimizing manual provisioning.

Table of Contents

1. Introduction

Software Defined Wide Area Network (SD-WAN), as described in [MEF70.1] and [MEF70.2], provides overlay connectivity services that optimize the transport of IP packets across one or more underlay networks by identifying traffic types and applying policies to determine forwarding behavior. Key characteristics of SD-WAN networks include:

- Transport Augmentation: an SD-WAN path can utilize different types of underlay networks, including private networks (with or without encryption) and public networks (requiring encryption).
- Direct Internet Breakout: Traffic from remote branch offices can directly access the internet, avoiding backhauling to corporate headquarters for centralized policy control.
- Policy-Based Traffic Steering: Traffic can be directed over specific overlay paths based on predefined conditions, such as matching one or multiple fields in the IP header, rather than solely relying on destination IP addresses [RFC9522]. For IPv6 [RFC8200], attributes like the Flow Label, source address, specific extension header fields, or a combination of these can be used. Additional details are available in Tables 7 and 8 of [MEF70.1].
- Performance-Based Forwarding: Traffic can be steered based on performance metrics (e.g., packet delay, loss, jitter), selecting the underlay path that meets or exceeds policy requirements.

This document outlines SD-WAN use cases and addresses the complexities of managing large-scale SD-WAN overlay networks, as described in [Net2Cloud-Problem]. It demonstrates how a BGP-based

control plane can efficiently manage these networks with minimal manual intervention.

It’s important to distinguish the BGP instance as the control plane for SD-WAN overlay from the BGP instances governing the underlay networks. The document assumes a secure communication channel between the SD-WAN controller and SD-WAN edges for exchanging control plane information.

The need for an RFC documenting SD-WAN use cases lies in ensuring standardization and interoperability. While BGP and IPsec are well-established technologies, their application to SD-WAN introduces challenges such as scalability, traffic segmentation, and multi-homing. This document consolidates best practices and defines guidelines to enable consistent implementations across diverse networks, optimizing existing protocols for SD-WAN scenarios rather than proposing new ones.

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 SD-WAN controller to manage SD-WAN overlay networks in this document. In the context of BGP-controlled SD-WAN, the SD-WAN controller functions as or is integrated with the BGP Route Reflector (RR).

Client service: A service (e.g., IP prefix or VLAN) attached to the client-facing interface of an SD-WAN edge node.

Client route: A BGP-advertised route originated by an SDWAN-Edge that represents the reachability of a client-facing service (e.g., IP prefix or VLAN) and includes associated path attributes used by the SDWAN-Controller for policy enforcement and forwarding decisions.

C-PE: SD-WAN Edge node, which can be Customer Premises Equipment (CPE) for customer-managed SD-WAN, or

Provider Edge (PE) for provider-managed SD-WAN services.

Homogeneous Encrypted SD-WAN: An SD-WAN network in which all traffic to/from the SD-WAN edges are carried by IPsec tunnels regardless of underlying networks. I.e., the client traffic is carried by IPsec tunnel even over MPLS private networks.

MP-NLRI: In this document, the term "MP-NLRI" serves as a concise reference for "MP_REACH_NLRI".

NSP: Network Service Provider.

PE: Provider Edge

SD-WAN Edge Node: A device, either physical or virtual, that participates in the SD-WAN overlay network. These nodes advertise client routes to the SD-WAN Controller (e.g., BGP RR).

SD-WAN: An overlay connectivity service that optimizes the transport of IP packets over one or more Underlay connectivity services by recognizing applications and determining forwarding behavior by applying policies to them. [MEF-70.1].

SD-WAN IPsec SA: IPsec Security Association between two WAN ports of the SD-WAN edge nodes or between two SD-WAN edge nodes.

SD-WAN over Hybrid Underlay Networks: SD-WAN over Hybrid Underlay Networks typically have edge nodes utilizing bandwidth resources from different types of underlay networks, some being private networks and others being public Internet.

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

Private VPN: refers to a VPN that is supported wholly by a single network service provider without using any elements of

the public Internet and without any traffic passing out of the immediate control of that service provider.

ZTP: Zero Touch Provisioning

3. SD-WAN Scenarios and Their Requirements

This section outlines the core requirements for SD-WAN overlay networks and introduces various SD-WAN scenarios. These scenarios serve as examples that are further explored in subsequent sections to illustrate how the BGP control plane is used to distribute reachability and policy information within SD-WAN overlay networks.

3.1. Requirements

3.1.1. Supporting SD-WAN Segmentation

"SD-WAN Segmentation" refers to policy-driven network partitioning, a common approach in SD-WAN deployment. An SD-WAN segment is essentially a virtual private network (SD-WAN VPN) consisting of a set of edge nodes interconnected by tunnels, such as IPsec tunnels and/or MPLS VPN tunnels.

This document assumes that SD-WAN VPN configuration on PE devices will, as with MPLS VPN [RFC4364] [RFC4659], make use of VRFs [RFC4364] [RFC4659]. Notably, a single SD-WAN VPN can be mapped to one or multiple virtual topologies governed by the SD-WAN controller’s policies.

When BGP is used for SD-WAN, the client route UPDATE is the same as MPLS VPN. The Route Target in the BGP Extended Community [RFC4360] can differentiate the routes belonging to different SD-WAN VPNs.

As SD-WAN is an overlay network arching over multiple types of networks, MPLS L2VPN[RFC4761] [RFC4762]/L3VPN[RFC4364] [RFC4659] or pure L2 underlay can continue using the VPN ID (Virtual Private Network Identifier), VN-ID (Virtual Network Identifier), or VLAN (Virtual LAN) in the data plane to differentiate packets belonging to different SD-WAN VPNs. For packets transported through an IPsec tunnel, additional encapsulation, such as GRE [RFC2784] or VxLAN [RFC7348], is needed to embed the SD-WAN VPN identifier inside the IPsec ESP header.

3.1.2. Client Service Requirement

The client service requirements describe the SD-WAN edge’s ports, also known as SD-WAN client interfaces, which connect the client network to the SD-WAN service.

The SD-WAN client interface should support IPv4 & IPv6 addresses as well as Ethernet in accordance with the [IEEE802.3] standard.

In [MEF 70.1], the "SD-WAN client interface" is called SD-WAN UNI (User Network Interface). Section 11 of [MEF 70.1] defines a comprehensive set of attributes for the SD-WAN UNI, detailing the expected behavior and requirements to enable seamless connectivity to the client network.

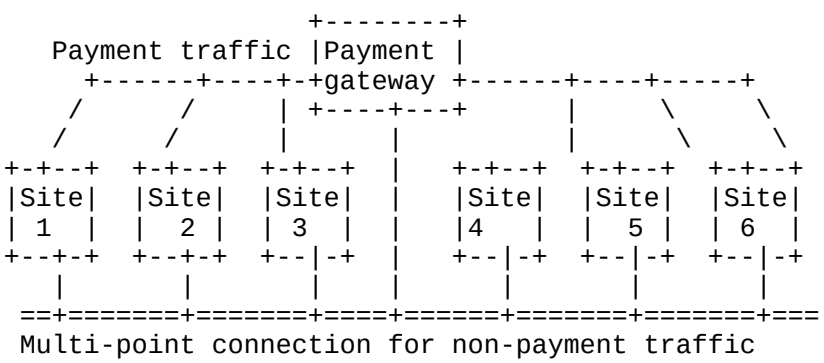
The client service at the SD-WAN edge must support the SD-WAN UNI service attributes outlined in Section 11 of [MEF 70.1].

3.1.3. SD-WAN Traffic Segmentation

SD-WAN Traffic Segmentation allows traffic to be separated based on business priorities, security requirements, and operational needs. This ensures that different user groups or services can operate within distinct topologies or follow tailored policies to meet specific business and security objectives.

For example, in a retail environment, traffic from point-of-sales (PoS) systems may require a different topology that is separate from other traffic. The PoS traffic is routed exclusively to the payment processing entity at a central hub site, while other types of traffic can be routed among all branches or remote sites.

In the figure below, traffic from the PoS system follows a tree topology (denoted as "----" in the figure below), whereas other traffic can follow a multipoint-to-multipoint topology (denoted as "===").



Another example is an enterprise that wants to isolate traffic by departments, ensuring each department having its own unique topology and policies. For instance, the HR department may need to access specific systems or resources that are not accessible by the engineering department. Similarly, contractors may have limited access to the enterprise resources.

3.1.4. Zero Touch Provisioning

SD-WAN Zero-Touch Provisioning (ZTP) is a network automation approach that enables the automatic provisioning and configuration of SD-WAN devices, such as routers and switches, at remote locations without requiring manual intervention. ZTP allows devices to be shipped with factory default settings; upon connection to the network, they automatically retrieve their configurations. ZTP for a remote SD-WAN edge usually includes the following steps:

- The SD-WAN edge’s customer information and unique device identifier (e.g., serial number, MAC address, or factory-assigned ID) are registered with the SD-WAN Central Controller.

- Upon power-up, the SD-WAN edge can establish the transport layer secure connection [BCP195] to its controller, whose URL (or IP address) and credential for connection request can be preconfigured on the edge device by the manufacture, external USB drive or secure Email given to the installer. The external USB method involves providing the installer with a pre-configured USB flash drive containing the necessary configuration files and settings for the SD-WAN device. The secure Email approach entails sending a secure email containing the configuration details for the SD-WAN device.

- The SD-WAN Controller authenticates the ZTP request from the remote SD-WAN edge with its configurations. Once the authentication is successful, it can designate a local network controller near the SD-WAN edge to pass down the initial configurations via the secure channel. The local network controller manages and monitors the communication policies for traffic to/from the edge node.

3.1.5. Constrained Propagation of SD-WAN Edge Properties

For an SD-WAN Edge to establish an IPsec tunnel to another edge and exchange the attached client routes, both edges need to know each other’s network properties, such as the IP addresses of the WAN ports, the edges’ loopback addresses, the attached client routes, the supported encryption methods, etc.

In many cases, an SD-WAN edge node is authorized to communicate with only a subset of other edge nodes. To maintain security and privacy, the property of an SD-WAN edge node must not be propagated to unauthorized peers. However, when a remote SD-WAN edge node powers up, it may lack the policies to determine which peers are authorized to communicate. Therefore, SD-WAN deployment needs to have a central point to distribute the properties of an SD-WAN edge node to its authorized peers.

BGP is well suited for this purpose. A Route-Reflector (RR) [RFC4456], integrated into the SD-WAN controller, enforces policies governing the communication among SD-WAN edges. The RR ensures that BGP UPDATE messages from an SD-WAN edge are propagated only to other edges within the same SD-WAN VPN.

An SD-WAN edge must use a secure channel, such as TLS (RFC5246) [RFC8446] or IPsec, to its designated RR for exchanging BGP UPDATE messages.

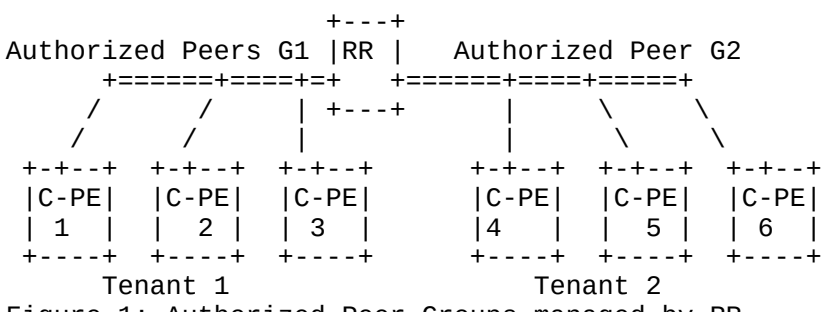


Figure 1: Authorized Peer Groups managed by RR

Tenant separation is achieved by the SD-WAN VPN identifiers represented in the control plane and data plane, respectively.

3.2. Scenario #1: Homogeneous Encrypted SD-WAN

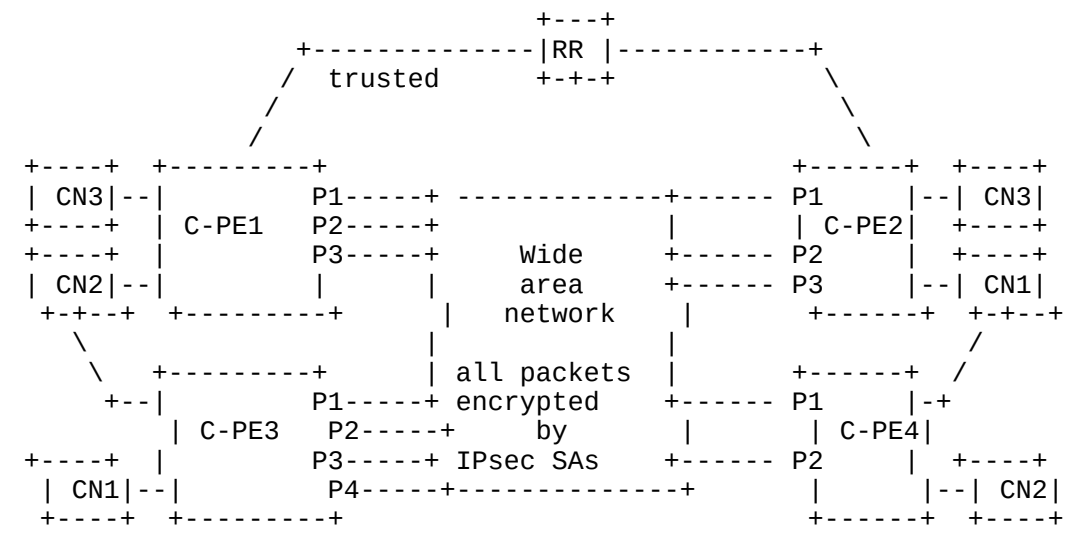
Homogeneous Encrypted SD-WAN refers to an SD-WAN network where edge nodes encrypt all client traffic destined to other edge nodes, regardless of whether the underlay is private or public.

Typical use cases for Homogeneous Encryption:

- A small branch office connecting to its headquarters via the Internet. All traffic to and from this small branch office must be encrypted, usually achieved by IPsec Tunnels [RFC6071].

- A retail store in a shopping mall may need to securely connect to its services hosted in one or more Cloud DCs via the Internet. A common method involves establishing IPsec SAs with the Cloud DC gateway to securely transport sensitive data to/from the store.

The granularity of the IPsec SAs for Homogeneous Encryption 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 subnet/tenant/site is encrypted.



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

Figure 2: Homogeneous Encrypted SD-WAN

A Homogeneous Encrypted SD-WAN shares certain similarities with traditional IPsec VPN. However, unlike IPsec VPNs, which are typically deployed in a point-to-point fashion among a limited number of nodes, SD-WAN networks can comprise a large number of edge nodes, all centrally managed by a controller responsible for configurations and policies across the network.

Existing private VPNs (e.g., MPLS based) can use Homogeneous Encrypted SD-WAN to extend over the public network to remote sites to which the VPN operator does not own or lease infrastructural connectivity.

3.3. Scenario #2: Differential Encrypted SD-WAN

Differential Encrypted SD-WAN refers to an SD-WAN network that utilizes hybrid underlays, combining private VPNs and the public Internet. In this model, traffic traversing the private VPN is forwarded natively without encryption, while traffic over the public Internet is encrypted for security. This approach balances performance and security. Since IPsec encryption requires significant processing power and traffic over the public Internet typically lacks the premium SLA (Service Level Agreement) provided by private VPNs-especially over long distances-current practice is to forward traffic over private VPNs without encryption, leveraging the inherent reliability and security of the private

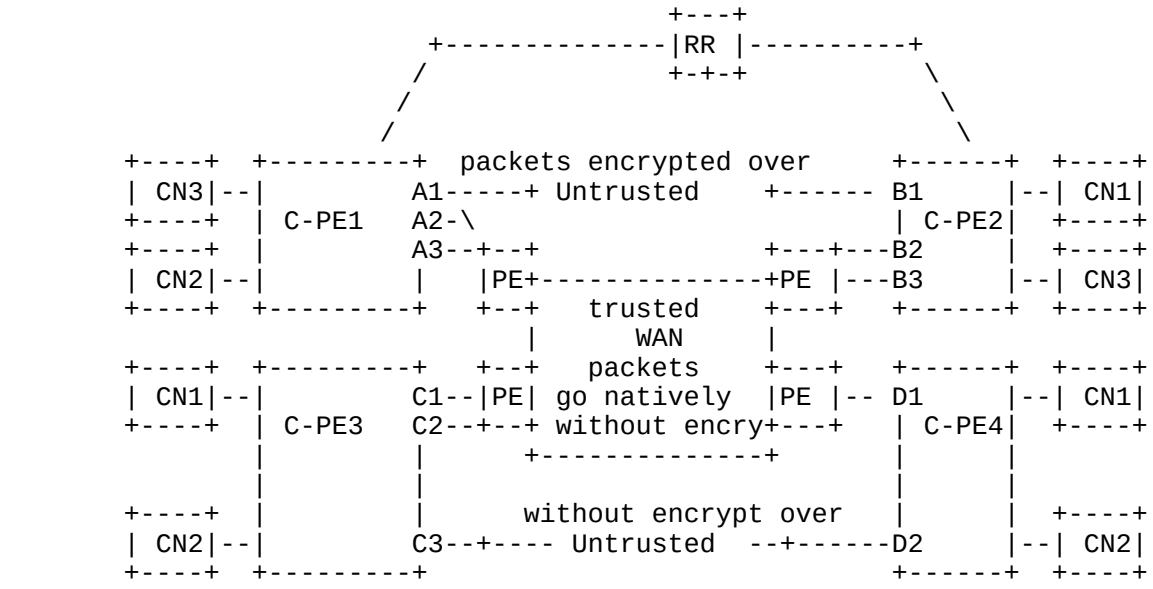
network. Meanwhile, encryption is applied only to traffic routed over the public Internet to ensure data confidentiality..

One C-PE might have the Internet-facing WAN ports managed by different NSPs with the WAN ports' addresses assigned by the corresponding NSPs. Clients may define specific policies to govern how traffic flows across the network, such as:

1) Certain flows can only be forwarded over private VPNs.
2) Certain flows can be forwarded over either private VPNs or the public Internet. When forwarded over the public Internet, the packets are encrypted.
3) Some flows, especially Internet-bound browsing ones, can be handed off to the Internet without further encryption.

For example, consider a flow traversing multiple segments, A<->B, B<->C, C<->D, has Policy 2) above. This flow can cross different underlays in different segments, such as over Private underlay between A<->B without encryption or over the public Internet between B<->C protected by an IPsec SA.

In the figure below, C-PE1 has two different types of interfaces: A1 to the Internet, and A2 & A3 to a private VPN. The WAN ports' addresses can be allocated by the service providers or dynamically assigned (e.g., by DHCP).



CN: Client Network

Figure 3: SD-WAN with Hybrid Underlays

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

3.4. Scenario #3: Private VPN PE based SD-WAN

Private VPN PE-Based SD-WAN refers to extending an existing VPN (e.g., EVPN [RFC7432] or IPVPN) by adding additional ports that face the public Internet to address increased bandwidth requirements between Provider Edge (PE) devices. This approach allows VPN service providers to augment their networks without immediately committing to building or leasing new infrastructure.

Key Characteristics of Private VPN PE-Based SD-WAN:

- For MPLS-based VPN, traffic between PEs uses MPLS encapsulation within IPsec tunnels egressing the Internet WAN ports, such as MPLS-in-IP or GRE-in-IPsec.

- The BGP RR remains connected to the PEs via the same trusted network as the original VPN, ensuring consistency in routing policies and security.

The main use case for Private VPN PE-Based SD-WAN is Temporary Bandwidth Expansion.

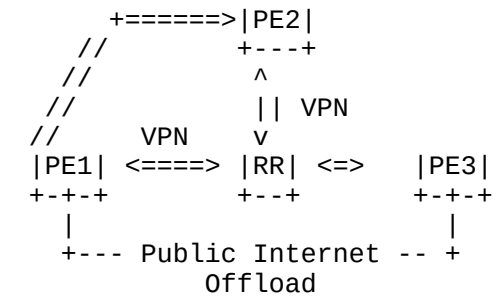


Figure 4: Additional Internet paths added to the VPN

For Ethernet-based client traffic, Private VPN PE based SD-WAN should support VLAN-based service interfaces (EVPN Instances), VLAN bundle service interfaces, or VLAN-Aware bundling service interfaces. EVPN service requirement as described in Section 3.1 of [RFC8388] are applicable to the SD-WAN Ethernet-based Client services. For IP-based client interfaces, L3VPN service requirements are applicable.

4. Provisioning Model

4.1. Client Service Provisioning Model

Provisioning of client-facing services in an SD-WAN network can leverage approaches similar to those used for VRFs (Virtual Routing and Forwarding) in MPLS based VPNs [RFC4364][RFC4659]. A client VPN can define communication policies by specifying BGP Route Targets for import and export. Alternatively, policy-based filtering using ACLs (Access Control List) can be employed to control which routes are allowed or denied for a given client VPN.

In scenarios where an SD-WAN edge node is dedicated to a single client with a single virtual network, all services attached to the client facing interface(s) on the edge node can be grouped into a single VRF. The RR can manage the policies for import/export policies for that VRF.

4.2. Policy Configuration

Policy configuration is a key characteristic of an SD-WAN service, enabling packets to be forwarded over multiple types of underlays based on predefined rules. Policies determines which underlay paths are allowed to carry specific flows, as outlined in Section 8 of [MEF70.1]. A flow is a collection of packets between the same source and destination pair that are subject to the same forwarding and policy decisions at the ingress SD-WAN edge node and are identified by the settings of one or more fields in the packet headers. For example, client-service-x can only be mapped to a MPLS topology, ensuring traffic alignment with business or security requirements.

4.3. IPsec Related Parameters Provisioning

IPsec-related parameter provision in an SD-WAN network involves the negotiation and distribution of cryptographic parameters required to establish IPsec tunnels among them. To streamline the configuration process, SD-WAN edge nodes can retrieve those parameters directly from the SD-WAN controller, reducing manual intervention and ensuring consistency.

In a BGP-controlled SD-WAN, BGP UPDATE messages can be extended to propagate IPsec-related attributes for each SD-WAN edge. This approach allows peers to receive and apply compatible cryptographic parameters distributed over a secure channel between the SDWAN-Edge and its BGP RR, thereby simplifying IPsec tunnel establishment and reducing reliance on traditional IKEv2 negotiation [RFC7296].

5. BGP Controlled SD-WAN

5.1. Rational for Using BGP as Control Plane for SD-WAN

In small SD-WAN networks with a modest number of nodes, traditional approaches such as the hub-and-spoke model, employing Next Hop Resolution Protocol (NHRP)[RFC2332] or a centralized hub

managing edge nodes, including the mapping of local and public addresses along with tunnel identifiers, has proven effective. However, for larger SD-WAN networks, with more than 100 nodes and encompassing diverse underlays, the conventional approach becomes increasingly complex, error-prone, and difficult to manage.

BGP offers several key advantages when used as the control plane for a large SD-WAN:

- Simplified peer authentication process:

  With a secure management channel established between each edge node and its RR, the RR can perform peer authentication on behalf of the edge node. The RR has policies on peer communication and the built-in capability to constrain the propagation of the BGP UPDATE messages to the authorized edge nodes only.

- Scalable IPsec tunnel management

  In networks with multiple IPsec tunnels between SD-WAN edges, BGP simplifies tunnel management by using the Tunnel Encapsulation Attribute specified in [RFC9012] to carry information that associates advertised client routes with specific tunnels.

  Unlike traditional IPsec VPN where IPsec tunnels between two edge nodes are treated as independent parallel links requiring duplicated control plane messages for load sharing.

- Simplified traffic selection configurations

  BGP can simplify the configuration of IPsec tunnel associations and related forwarding policies. By leveraging Route Targets to identify SD-WAN VPN membership, administrators can apply import/export policies that control the distribution of client routes. These route attributes, in turn, inform the local configuration of IPsec traffic selectors at each SDWAN-Edge.

- Centralized Management and Security
  When the BGP RR is integrated with the SD-WAN controller, it supports a centralized model for managing route distribution policies. The RR ensures that BGP UPDATE messages are distributed only to authorized SD-WAN Edges based on preconfigured policies, thereby reducing control-plane

complexity and limiting exposure compared to decentralized architectures.

In summary, BGP combines scalability, robust policy enforcement, interoperability, and centralized security, making it an ideal choice for managing SD-WAN overlay networks, particularly as they grow in size and complexity.

5.2. BGP Scenario for Homogeneous Encrypted SD-WAN

In a BGP-controlled Homogeneous Encrypted SD-WAN, an SD-WAN Edge (i.e., C-PE) can advertise both its attached client routes and associated IPsec tunnel parameters using BGP UPDATE messages, potentially within in a single message that includes the Tunnel Encapsulation Attribute.

For example, in the figure below, the BGP UPDATE message from C-PE2 to RR can have the client routes encoded in the MP-NLRI Path Attribute and the IPsec Tunnel associated parameters encoded in the Tunnel Encapsulation Attribute [RFC9012].

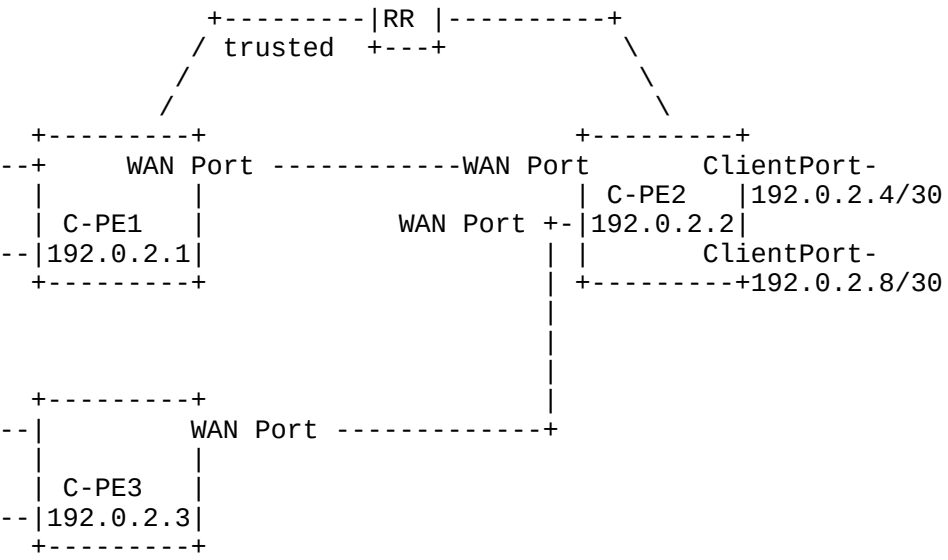


Figure 5: Homogeneous Encrypted SD-WAN

In scenarios where C-PE2 does not have a policy specifying the authorized peers for specific client routes, the RR takes the

responsibility for ensuring that BGP UPDATE for these client routes are propagated only to other authorized SD-WAN edges.

5.3. BGP Scenario for Differential Encrypted SD-WAN

In this scenario, client services may have distinct forwarding requirements based on business or network policies. Some client services can be routed through any WAN ports of the edge node, while others must be routed through specific WAN ports (such as only MPLS VPN). To address these requirements, the BGP speaker employs two distinct BGP UPDATE messages:

- Update 1: Client Route Advertisement for advertising the prefixes of client services attached to the client facing interfaces. The Color (Section 8 of [RFC9012]) is used to associate each client service with the corresponding WAN ports for the desired underlay paths.
- Update 2: Underlay WAN Port Advertisement, which conveys information about the underlay WAN facing interfaces of an SD-WAN Edge, including attributes such as IPsec SA parameters, MPLS label stacks, and other relevant attributes. These attributes are carried in the BGP Tunnel Encapsulation Attribute, along with associated Color values that allow BGP receivers to correlate the WAN facing interfaces with the client routes advertised in Update 1.

This dual-update approach provides flexibility and efficiency in managing IPsec tunnels terminated at the WAN ports of SD-WAN edge nodes. By decoupling client route advertisements from IPsec tunnel attributes, it accommodates the differing update frequencies between these components-for example, client route changes may occur independently of dynamic IPsec parameters such as key values. Additionally, multiple client services can share a single IPsec SA, optimizing resource usage and reducing control-plane overhead.

BGP receivers associate the two UPDATE messages using the common loopback address of the SDWAN-Edge (e.g., C-PE2). UPDATE 1 advertises client routes with the next-hop set to C-PE2’s loopback address. UPDATE 2 advertises underlay WAN port information using an NLRI that contains the same loopback address, along with the Tunnel Encapsulation Attribute conveying IPsec parameters and WAN port properties. BGP receivers use the common loopback address to match the next-hop in UPDATE 1 with the NLRI in UPDATE 2. This enables recursive resolution, as specified in [RFC9012], allowing

client traffic to be forwarded based on the underlay characteristics defined in UPDATE 2.

5.4. BGP Scenario for Flow-Based Segmentation

In a flow-based segmentation scenario, as described in Section 3.1.3, a service flow is identified by specific fields in the packet's IP header, such as source/destination IP addresses, port numbers, or protocol types. Flow-based segmentation ensures that traffic for a particular service flow is directed only to authorized nodes or paths, meeting security and policy requirements.

This can be achieved by constraining the propagation of BGP UPDATE messages to nodes that meet the criteria of the service flow. For instance, to enforce communication exclusively between the Payment Application in branch locations and the Payment Gateway, as depicted in Figure 6, the following BGP UPDATE messages can be advertised:

BGP UPDATE #1a: Propagated only to the Payment Gateway node for a point-to-point (P2P) topology between the Payment Application and the Payment Gateway.

BGP UPDATE #1b: Propagated to C-PE1 and C-PE3 for other prefixes that can be reached by these edge nodes.

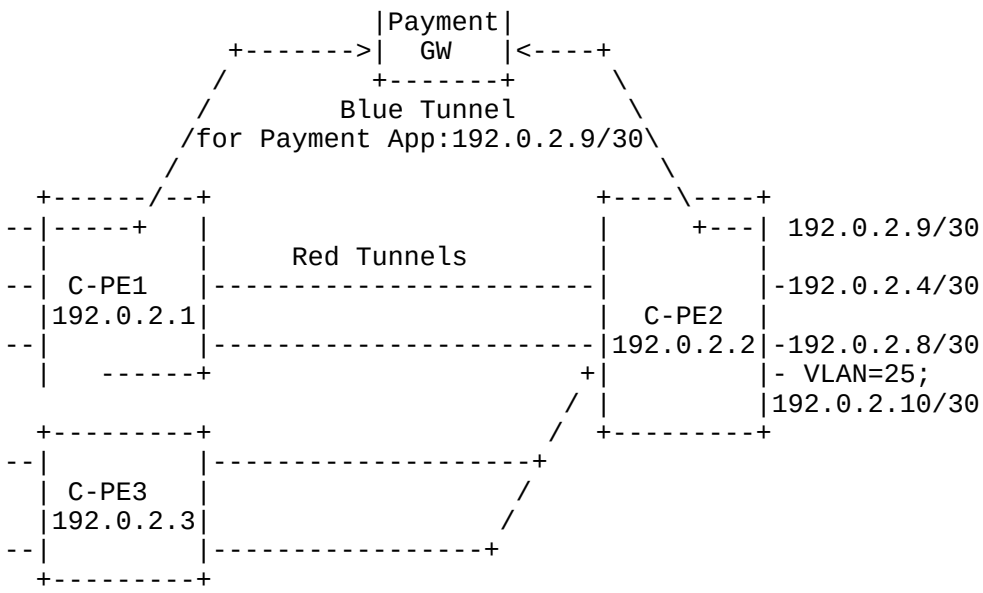


Figure 6: Flow Based SD-WAN Segmentation

6. SD-WAN Forwarding Model

This section describes how client traffic is forwarded in a BGP Controlled SD-WAN for the use cases described in Section 3.

The forwarding procedures described in Section 6 of [RFC8388] are applicable for the SD-WAN client traffic. Similar to the BGP-based VPN/EVPN client routes UPDATE message, Route Targets can be used to distinguish routes from different clients.

6.1. Forwarding Model for Homogeneous Encrypted SD-WAN

6.1.1. Network and Service Startup Procedures

In the Homogeneous Encrypted SD-WAN scenario, two IPsec SAs are required to secure bidirectional traffic between two C-PE nodes (or their client-facing interfaces), since each SA protects traffic in only one direction.

For example, in the full mesh scenario in Figure 2 of Section 3.2, where client CN2 is attached to C-PE1, C-PE3, and C-PE4, six uni-directional IPsec SAs must be established: C-PE1 <-> C-PE3; C-PE1 <-> C-PE4; C-PE3 <-> C-PE4.

SD-WAN services to clients can be IP-based or Ethernet-based. For IP-based services, an SD-WAN edge can learn client addresses from the client-facing interfaces via OSPF, RIP, BGP, or static configuration. For Layer-2 services, the EVPN parameters, such as the ESI (Ethernet Segment Identifier), EVI (Ethernet Virtual Instance), and CE-VID (Customer Edge Virtual Instance Identifier) to EVI mapping, can be configured as described in [RFC8388].

Instead of running IGP within each IPsec tunnel, as is common in traditional IPsec VPN, the BGP RR can propagate the client route UPDATE messages to authorized SD-WAN Edges based on configured policies. The SDWAN-Edges use BGP attributes-such as the Tunnel Encapsulation Attribute and associated Color values-to associate received client routes with the appropriate IPsec Security Associations (SAs), thereby eliminating the need for manual configuration of tunnel endpoints and service policies on each edge node.

6.1.2. Packet Walk-Through

For unicast packets forwarding:

An IPsec SA terminated at a C-PE node can carry traffic for multiple client services. Packets to and from these services are encapsulated in an inner tunnel, such as GRE or VXLAN. Different client traffic can be distinguished using a unique key or identifier in the inner encapsulation header. This inner tunnel is further encapsulated with an outer IP header, where the source and destination addresses are the loopback addresses of the C-PE nodes, and the protocol field is typically set to ESP (50).

C-PE Node-based IPsec tunnel is inherently protected when the C-PE has multiple WAN ports to different underlay paths. As shown in Figure 2, when one of the underlay paths fails, the IPsec tunnel can continue operating by rerouting traffic through an alternate WAN port.

When a C-PE receives an IPsec encrypted packet from its WAN ports, it decrypts the packet and forwards the inner packet to the client facing interface based on the inner packet’s destination address.

For multicast packets forwarding:

IPsec was created to be a security protocol between two and only two devices, so multicast service using IPsec is problematic. An IPsec peer encrypts a packet so that only one other IPsec peer

can successfully perform the de-encryption. A straightforward way to forward a multicast packet for the Homogeneous Encrypted SD-WAN is to encapsulate the multicast packet in separate unicast IPsec SA tunnels. More optimized forwarding multicast packets for the Homogeneous Encrypted SD-WAN is out of the scope of this document.

6.2. Forwarding Model for Hybrid Underlay SD-WAN

In this scenario, as shown in Figure 3 of Section 3.3, traffic forwarded over the trusted VPN paths can be native (i.e., unencrypted). The traffic forwarded over untrusted networks need to be protected by IPsec SA.

6.2.1. Network and Service Startup Procedures

Infrastructure setup: The proper MPLS infrastructure must be configured among the edge nodes, i.e., the C-PE1/C-PE2/C-PE3/C-PE4 of Figure 3. The IPsec SA between wAN ports or nodes must be set up as well. IPsec SA related attributes on edge nodes can be distributed by BGP UPDATE messages as described in Section 5.

There could be policies governing how flows can be forwarded, as specified by [MEF70.1]. For example, "Private-only" indicates that the flows can only traverse the MPLS VPN underlay paths.

6.2.2. Packet Walk-Through

Unicast packets forwarding:

When C-PE-a in Figure 7 receives a packet from a client facing interface, the forwarding decision depends on the flow’s routing policy. If a packet belonging to a flow that must be forwarded over the MPLS VPN, the forwarding processing is the same as the MPLS VPN. Otherwise, C-PE-a can select the least cost path, including the previously established MPLS paths and IPsec Tunnels, to forward the packet. Packets forwarded over the trusted MPLS VPN do not require additional encryption, while those sent over untrusted networks must be encrypted by IPsec SA.

For a c-PE with multiple WAN ports provided by different NSPs, separate IPsec SAs can be established for the WAN ports. In this case, the C-PE have multiple IPsec tunnels in addition to the MPLS path to choose from to forward the packets from the client facing interfaces.

If the IPsec SA is chosen, the packet is encapsulated by the IPsec header and encrypted by the IPsec SA before forwarding it to the WAN.

Packets received over MPLS paths are processed as in standard MPLS VPNs. For packets encrypted with IPsec received from WAN ports, the C-PE decrypts and decapsulates the inner payload before forwarding it according to the local forwarding table. To protect against potential attacks, traffic received through Internet-facing WAN ports must undergo anti-DDoS mechanisms, which are beyond the scope of this document. Additionally, the control plane must avoid learning routes from Internet-facing WAN ports to ensure network integrity.

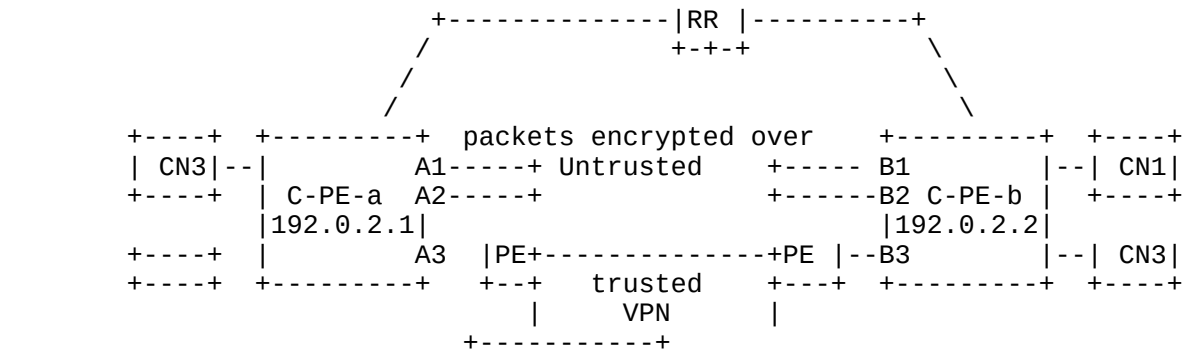


Figure 7: Over hybrid SD-WAN

Multicast packets forwarding:

For multicast traffic, MPLS multicast [RFC6513, RFC6514, or RFC7988] can be utilized to forward multicast traffic across the network.

If IPsec tunnels are used for multicast traffic, the packet must be encapsulated and encrypted separately for each destination, creating multiple unicast IPsec tunnels to deliver the multicast packet to all intended recipients.

6.3. Forwarding Model for PE based SD-WAN

6.3.1. Network and Service Startup Procedures

In this scenario, all PEs have secure interfaces facing the clients and facing the MPLS backbone. Some PEs have additional interfaces to the untrusted public Internet which are for offloading low priority traffic when the MPLS paths get congested.

The PEs are already connected to their RRs, and the configurations for the clients and policies are already established.

6.3.2. Packet Walk-Through

When offloading MPLS packets to the Internet path, each MPLS packet is encapsulated by an outer IP header as MPLS-in-IP or MPLS-in-GRE [RFC4023]. The outer IP address can be an interface address or the PE's loopback address.

When IPsec Tunnel mode is used to protect an MPLS-in-IP packet, the entire MPLS-in-IP packet is placed after the IPsec tunnel header. In IPsec transport mode, the MPLS-in-IP packet's IP header becomes the outer IP header of the IPsec packet, followed by an IPsec header, and then followed by the MPLS label stack. The IPsec header must set the payload type to MPLS by using the IP protocol number specified in section 3 of [RFC4023]. For the MPLS-in-GRE packets protected by IPsec Transport Mode, the GRE header follows the IPsec header.

The IPsec SA's endpoints should not be the client-facing interface addresses unless the traffic to/from those clients always goes through the IPsec SA even when the MPLS backbone has enough capacity to transport the traffic.

When the PEs' Internet-facing ports are behind the NAT [RFC3715], additional measures are necessary to support NAT traversal. In this Case, an outer UDP field is added to the encrypted payload [RFC3948]. Three specific ports and protocols must remain open on the PEs: UDP port 4500 (used for NAT traversal), UDP port 500 (used for IKE), and IP protocol 50 (ESP). The IPsec IKE (Internet Key Exchange) sessions between PEs navigate NAT environment using the mechanisms outlined in [RFC3947].

When a packet is received from a client facing interface, it is initially processed according to the MPLS VPN forwarding rules. If the MPLS backbone path to the destination is congested, the packet is encapsulated as an MPLS-in-IP packet and encrypted using the IPsec tunnel to the target PE. Conversely, when a packet is received from an Internet-facing WAN port, it is decrypted, and the inner MPLS payload is extracted and forwarded to the MPLS VPN engine for further processing.

Same as Scenario #2, the additional anti-DDoS mechanism must be enabled to prevent potential attacks from the Internet-facing port. Control Plane should not learn routes from the Internet-facing WAN ports.

7. Manageability Considerations

A BGP-controlled SD-WAN uses RR to propagate client routes and underlay tunnel properties among authorized SD-WAN edges. Since the RR is configured with policies that identify authorized peers, the peer-wise IPsec IKE (Internet Key Exchange) authentication process is significantly simplified.

8. Security Considerations

In a BGP-controlled SD-WAN network, secure operation replies in part on the correct configuration and behavior of the RR, which acts as the central distribution point for BGP routing information. RR applies preconfigured routing policies to control the propagation of BGP UPDATE messages to authorized SD-WAN Edges, help minimizing the risk of unintended route exposure or unauthorized communication.

The security model for the SD-WAN described in this document is based on the following principles:

1) Centralized Control: The RR governs all routing and policy decisions. This centralized architecture simplifies security management compared to distributed models, as it limits the potential attack surface to a smaller, more controlled set of components.
2) Secure Communication Channels: All communication between SD-WAN edge nodes and the RR must occur over a secure channel, such as TLS or IPsec, to ensure the confidentiality and integrity of BGP UPDATE messages.
3) Policy Enforcement: The RR is responsible for enforcing policies that restrict the propagation of edge node properties and routing updates to only authorized peers. This prevents sensitive information from being exposed to unauthorized nodes.
4) Mitigation of Internet-Facing Risks: In scenarios where SD-WAN edge nodes include Internet-facing WAN ports, additional measures must be taken to mitigate security risks:
   - Anti-DDoS mechanisms must be enabled to protect against potential attacks on Internet-facing ports.
   - The control plane must avoid learning routes from Internet-facing WAN ports to prevent unauthorized traffic from being injected into the SD-WAN.

By concentrating the security within the RR and using secure communication channels, the SD-WAN network achieves consistent enforcement of security policies and reduces the likelihood of misconfigurations at individual edge nodes. However, the robustness of this security model depends critically on the proper configuration and ongoing maintenance of the RR. Operators must ensure that the RR itself is adequately protected against compromise or misconfiguration, as its failure or exploitation could impact the entire network.

This model emphasizes simplicity and efficiency, leveraging centralized governance to mitigate risks while ensuring scalability and interoperability of the SD-WAN.

9. IANA Considerations

No Action is needed.

10. References

10.1. Normative References

[BCP195] Consists of RFC8996 and RFC9325.

[RFC2332] J. Luciani, et al, "NBMA Next Hop Resolution Protocol (NHRP)", RFC2332, April 1998.

[RFC2784] D. Farinacci, et al, "Generic Routing Encapsulation (GRE)" RFC2784, March 2000.

[RFC3715] B. Aboba, W. Dixon, "IPsec-Network Address Translation (NAT) Compatibility Requirements", March 2004.

[RFC3947] T. Kivinen, et al, "Negotiation of NAT Traversal in the IKE", Jan. 2005.

[RFC3948] A. Huttunen, et al, "UDP Encapsulation of IPsec ESP Packets", Jan 2005.

[RFC4023] T. Worster, Y. Rekhter, E. Rosen, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", March 2005.

[RFC4360] S. Sangli, et al, "BGP Extended Communities Attribute", RFC4360, Feb. 2006.

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

[RFC4456] T. Bates, E. Chen, R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", April 2006.

[RFC4659] J. De clercq, et al, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC4659, Sept 2006.

[RFC4761] K. Kompella and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC4761, Jan. 2007.

[RFC4762] M. Lasserre and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC4762, Jan. 2007.

[RFC6071] S. Frankel, S. Krishan, "IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap", Feb 2011.

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

[RFC7348] M. Mahalingam, et al, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC7348, Aug 2014.

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

[RFC8200] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification". July 2017.

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

[RFC8388] J. Rabadan, et al, "Usage and Applicability of BGP MPLS-Based Ethernet VPN", RFC8388, May 2018.

[RFC9012] K.Patel, et al "The BGP Tunnel Encapsulation Attribute", RFC9012, April 2021.

[RFC9522] A. Farrel, "Overview and Principles of Internet Traffic Engineering", RFC9522, Jan. 2024.

10.2. Informative References

[Net2Cloud-Problem] L. Dunbar and A. Malis, "Dynamic Networks to Hybrid Cloud DCs: Problems and Mitigation Practices", draft-ietf-rtgwg-net2cloud-problem-statement-41, April. 2024.

[IEEE802.3] "IEEE Standard for Ethernet" by The Institute of Electrical and Electronics Engineers (IEEE) 802.3.

[MEF70.1] SD-WAN Service Attributes and Service Framework, https://www.mef.net/resources/mef-70-1-sd-wan-service-attributes-and-service-framework/. Nov 2021.

[MEF70.2] "SD-WAN Service Attributes and Service Framework" by MEF, https://www.mef.net/resources/mef-70-2-sd-wan-service-attributes-and-service-framework/. Oct 2023.

11. Acknowledgments

Acknowledgements to Gunter van de Velde, Andrew Alston, Adrian Farrel, Jim Guichard, Joel Halpern, John Scudder, Darren Dukes, Andy Malis, Donald Eastlake, Stephen Farrell, and Victo Sheng 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

Ali Sajassi
Cisco
Email: sajassi@cisco.com

John Drake
Independent
Email: je_drake@yahoo.com

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

Sue Hares
Email: shares@ndzh.com

Contributor's Addresses

David Carrel
Graphiant
Email: carrel@graphiant.com

Ayan Banerjee
Cisco
Email: ayabaner@cisco.com