

Network Working Group
Internet Draft
Intended status: Informational
Expires: May 1, 2020

L. Dunbar
Futurewei
A. Malis
Independent
C. Jacquenet
Orange
November 1, 2019

Gap Analysis of Dynamic Networks to Hybrid Cloud DCs
draft-ietf-rtgwg-net2cloud-gap-analysis-03

Abstract

This document analyzes the technological gaps when using SDWAN to dynamically interconnect workloads and applications hosted in various 3 party cloud data centers.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

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

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

This Internet-Draft will expire on May 1, 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.....	2
2.	Conventions used in this document.....	3
3.	Gap Analysis of C-PEs WAN Port Management.....	4
4.	Aggregating VPN paths and Internet paths.....	6
4.1.	Key Control Plane Components of SDWAN Overlay.....	7
4.2.	Using BGP UPDATE Messages.....	8
4.3.	SECURE-L3VPN/EVPN.....	9
4.4.	Preventing attacks from Internet-facing ports.....	10
5.	C-PEs not directly connected to VPN PEs.....	11
5.1.	Floating PEs to connect to Remote CPEs.....	13
5.2.	NAT Traversal.....	13
5.3.	Complexity of using BGP between PEs and remote CPEs via Internet.....	13
5.4.	Designated Forwarder to the remote edges.....	14
5.5.	Traffic Path Management.....	15
6.	Manageability Considerations.....	15
7.	Security Considerations.....	15
8.	IANA Considerations.....	16
9.	References.....	16
9.1.	Normative References.....	16
9.2.	Informative References.....	16
10.	Acknowledgments.....	17

[1. Introduction](#)

[Net2Cloud-Problem] describes the problems that enterprises face today in transitioning their IT infrastructure to support digital

economy, such as connecting enterprises' branch offices to dynamic workloads in different Cloud DCs.

This document analyzes the technological gaps to interconnect dynamic workloads & apps hosted in cloud data centers that the enterprise's VPN service provider may not own/operate or may be unable to provide the enterprise with the required connectivity to access such locations. When VPN service providers have insufficient bandwidth to reach a location, SDWAN techniques can be used to aggregate bandwidth of multiple networks, such as MPLS VPNs or the Public Internet to achieve better performance. This document primarily focuses on the technological gaps raised by using SDWAN techniques to connect enterprise premises to cloud data centers operated by third parties.

For the sake of readability, a SDWAN edge, a SDWAN endpoint, C-PE, or CPE are used interchangeably throughout this document. However, each term has some minor emphasis, especially when used in other related documents:

- . SDWAN Edge: could include multiple devices (virtual or physical);
- . SDWAN endpoint: to refer to a WAN port of SDWAN devices or a single SDWAN device;
- . C-PE: more for provider owned SDWAN edge, e.g. for SECURE-EVPN's PE based VPN, when PE is the edge node of SDWAN;
- . CPE: more for enterprise owned SDWAN edge.

2. Conventions used in this document

Cloud DC: Third party Data Centers that usually host applications and workload owned by different organizations or tenants.

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

CPE-Based VPN: Virtual Private Network designed and deployed from CPEs. This is to differentiate from most commonly used PE-based VPNs as in [RFC 4364](#).

OnPrem: On Premises data centers and branch offices

SDWAN: Software Defined Wide Area Network, "SDWAN" refers to the solutions of pooling WAN bandwidth from multiple underlay networks to get better WAN bandwidth management, visibility & control. When the underlay is a private network, traffic may be forwarded without any additional encryption; when the underlay networks are public, such as the Internet, some traffic needs to be encrypted when passing through (depending on user-provided policies).

3. Gap Analysis of C-Pes WAN Port Management

One of the key characteristics of the networks that interconnect workloads in Hybrid Cloud DCs is that those networks' edges can have WAN ports facing networks provided by different ISPs, some can be untrusted public internet, some can be MPLS VPN, some can be Cloud internal networks, some can be others.

If an edge node only has one single WAN port facing untrusted network, then all sensitive data to/from this edge have to be encrypted, usually by IPsec tunnels which can be terminated at the single WAN port address or at the edge node's internal address if it is routable in the wide area network.

If an edge node has multiple WAN ports with some facing private VPN and some facing public untrusted network, sensitive data can be forwarded via ports facing VPN natively without encryption and via ports facing public network with encryption. To achieve this flexibility, it is necessary to have the IPsec tunnels terminated at the WAN ports facing the public networks.

In order to establish pair-wise secure encrypted connection among those WAN ports, it is necessary for peers to be informed of the WAN port properties.

Some of those overlay networks (a.k.a. SDWAN in the context of this document) use the modified NHRP protocol [[RFC2332](#)] to register WAN ports of the edges with their "Controller" (or NHRP server), which then map a private VPN address to a public IP address of the destination node/port. DSVPN [[DSVPN](#)] or DMVPN [[DMVPN](#)] are used to establish tunnels between WAN ports of SDWAN edge nodes.

NHRP was originally intended for ATM address resolution, and as a result, it misses many attributes that are necessary for dynamic endpoint C-PE registration to the controller, such as:

- Interworking with the MPLS VPN control plane. An overlay (SDWAN) edge can have some ports facing the MPLS VPN network over which packets can be forwarded without any encryption and some ports facing the public Internet over which sensitive traffic needs to be encrypted before being sent.
- Scalability: NHRP/DSVPN/DMVPN works fine with small numbers of edge nodes. When a network has more than 100 nodes, these protocols do not scale well.
- NHRP does not have the IPsec attributes, which are needed for peers to build Security Associations over the public internet.
- NHRP messages do not have any field to encode the C-PE supported encapsulation types, such as IPsec-GRE or IPsec-VxLAN.
- NHRP messages do not have any field to encode C-PE Location identifiers, such as Site Identifier, System ID, and/or Port ID.
- NHRP messages do not have any field to describe the gateway(s) to which the C-PE is attached. When a C-PE is instantiated in a Cloud DC, it is desirable for C-PE's owner to be informed of how/where the C-PE is attached.
- NHRP messages do not have any field to describe C-PE's NAT properties if the C-PE is using private addresses, such as the NAT type, Private address, Public address, Private port, Public port, etc.

[BGP-SDWAN-PORT] describes how SDWAN edge nodes use BGP to register their WAN ports properties to the SDWAN controller, which then propagates the information to other SDWAN edge nodes that are authenticated and authorized to communicate with them.

4. Aggregating VPN paths and Internet paths

Most likely, enterprises (especially the largest ones) already have their C-PEs interconnected by providers VPNs, such as EVPN, L2VPN, or L3VPN, which can be PE-based or CPE-based. The commonly used PE-based VPNs have C-PE directly attached to PEs, therefore the communication between C-PEs and PEs is considered as secure. MP-BGP is used to learn & distribute routes among C-PEs, even though sometimes routes among C-PEs are statically configured on the C-PEs.

For enterprises already interconnected by VPNs, it is desirable to aggregate the bandwidth among VPN paths and Internet paths by C-PEs adding additional ports facing public internet. Under this scenario which is referred to as SDWAN Overlay throughout this document, it is necessary for the C-PEs to use a protocol to register their WAN port properties with their SDWAN Controller(s). The information is needed for the establishment and the maintenance of Port-based IPsec SA associations among relevant C-PEs.

When using NHRP for registration purposes, C-PEs need to run two separate control planes: EVPN&BGP for CPE-based VPNs, and NHRP & DSVPN/DMVPN for ports connected to the Internet. Two separate control planes not only add complexity to C-PEs, but also increase operational cost.

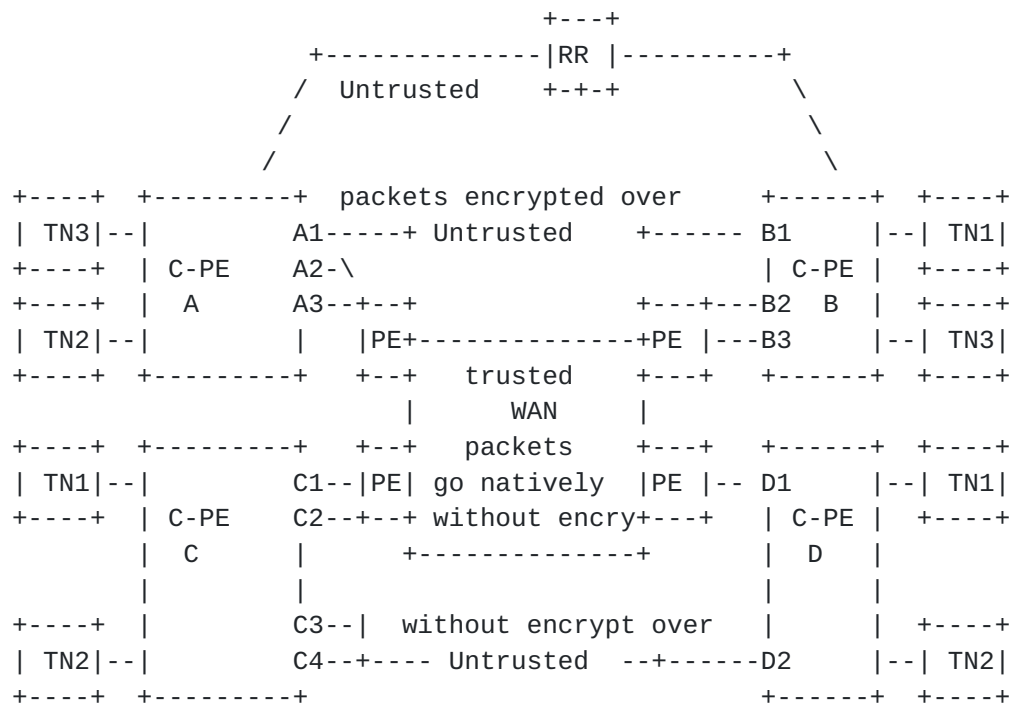


Figure 1: CPEs interconnected by VPN paths and Internet Paths

4.1. Key Control Plane Components of SDWAN Overlay

As described in [[BGP-SDWAN-Usage](#)], the SDWAN Overlay Control Plane has three distinct properties:

- SDWAN node's WAN Port Property registration to the SDWAN Controller.
 - o To inform the SDWAN controller and authorized peers of the WAN port properties of the C-PE [SDWAN-Port]. When the WAN ports are assigned private addresses, this step can register the type of NAT that translates private addresses into public ones.
- Controller facilitated IPsec SA management and NAT information distribution
 - o It is for SDWAN controller to facilitate or manage the IPsec configuration and peer authentication for all IPsec tunnels terminated at the SDWAN nodes.
- Establishing and Managing the topology and reachability for services attached to the client ports of SDWAN nodes.

- o This is for the overlay layer's route distribution, so that a C-PE can populate its overlay routing table with entries that identify the next hop for reaching a specific route/service attached to remote nodes. [[SECURE-EVPN](#)] describes EVPN and other options.

4.2. Using BGP UPDATE Messages

[[Tunnel-Encap](#)] describe the BGP UPDATE Tunnel Path Attribute that advertise endpoints' tunnel encapsulation capability for the respective attached client routes encoded in the MP-NLRI Path Attribute. The receivers of the BGP UPDATE can use any of the supported encapsulations encoded in the Tunnel Path Attribute for the routes encoded in the MP-NLRI Path Attribute.

Here are some of the gaps using [[Tunnel-Encap](#)] to distribute SDWAN Edge WAN port properties:

- [[Tunnel-Encap](#)] doesn't yet have the encoding to describe the NAT information for WAN ports that have private addresses. The NAT information needs to be propagated to the trusted peers via Controllers, such as the virtual C-PEs instantiated in public Cloud DCs.
- It is not easy using the current mechanism in [[Tunnel-Encap](#)] to exchange IPsec SA specific parameters independently from advertising the attached clients' routes, even after adding a new IPsec tunnel type.
[[Tunnel-Encap](#)] requires all tunnels updates are associated with routes. There can be many client routes associated with the SDWAN IPsec tunnel between two C-PEs' WAN ports; the corresponding destination prefixes (as announced by the aforementioned routes) may also be reached through the VPN underlay without any encryption.

The establishment of an IPsec tunnel can fail, such as due to the two endpoints supporting different encryption algorithms or other reasons. There can be multiple negotiations messages for the IPsec SA parameters between two end points. That is why IPsec SA association establishment between end points is independent from the policies on mapping routes to specific IPsec SA.

If C-PEs need to establish WAN Port based IPsec SA, the information encoded in Tunnel Path Attribute should only apply to the WAN ports and should be independent of the clients' routes.

In addition, the SDWAN Tunnel (IPsec SA) may need to be established before clients' routes are attached.

- C-PEs tend to communicate with a subset of the other C-PEs, not all the C-PEs need to be connected through a mesh topology. Therefore, the distribution of the SDWAN Overlay Edge WAN ports information need be be scoped to the authorized peers.

4.3. SECURE-L3VPN/EVPN

[SECURE-L3VPN] describes how to extend the BGP/MPLS VPN [[RFC4364](#)] capabilities to allow some PEs to connect to other PEs via public networks. [[SECURE-L3VPN](#)] introduces the concept of Red Interface & Black Interface used by PEs, where the RED interfaces are used to forward traffic into the VPN, and the Black Interfaces are used between WAN ports through which only IPsec-protected packets are forwarded to the Internet or to other backbone network thereby eliminating the need for MPLS transport in the backbone.

[SECURE-L3VPN] assumes PEs using MPLS over IPsec when sending traffic through the Black Interfaces.

[SECURE-EVPN] describes a solution where point-to-multipoint BGP signaling is used in the control plane for the SDWAN Scenario #1 described in [[BGP-SDWAN-Usage](#)]. It relies upon a BGP cluster design 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.

Both [[SECURE-L3VPN](#)] and [[SECURE-EVPN](#)] are useful, however, they both miss the aspects of aggregating VPN and Internet underlays. In summary:

- Both documents assume a client traffic is either forwarded all encrypted through an IPsec tunnel, or not encrypted at all through a different tunnel regardless which WAN ports the traffic egress the PEs towards WAN. In SDWAN, one client traffic can be forwarded encrypted at one time through a WAN port towards untrusted network

and be forwarded unencrypted at different time through a WAN port to MPLS VPN.

- The [\[SECURE-L3VPN\]](#) assumes that a CPE "registers" with the RR. However, it does not say how. It assumes that the remote CPEs are pre-configured with the IPsec SA manually. In SDWAN, Zero Touch Provisioning is expected. Manual configuration is not an option, especially for the edge devices that are deployed in faraway places.
- The [\[SECURE-L3VPN\]](#) assumes that C-PEs and RR are connected via an IPsec tunnel. Missing TLS/DTLS. The following assumption by [\[SECURE-L3VPN\]](#) becomes invalid for SDWAN environment where automatic synchronization of IPsec SA between C-PEs and RR is needed:
 - A CPE must also be provisioned with whatever additional information is needed in order to set up an IPsec SA with each of the red RRs
- IPsec requires periodic refreshment of the keys. The draft does not provide any information about how to synchronize the refreshment among multiple nodes.
- IPsec usually sends configuration parameters to two endpoints only and lets these endpoints negotiate the key. The [\[SECURE-L3VPN\]](#) assumes that the RR is responsible for creating/managing the key for all endpoints. When one endpoint is compromised, all other connections will be impacted.

[4.4.](#) Preventing attacks from Internet-facing ports

When C-PEs have Internet-facing ports, additional security risks are raised.

To mitigate security risks, in addition to requiring Anti-DDoS features on C-PEs, it is necessary for C-PEs to support means to determine whether traffic sent by remote peers is legitimate to prevent spoofing attacks.

5. C-PEs not directly connected to VPN PEs

Because of the ephemeral property of the selected Cloud DCs for specific workloads/Apps, an enterprise or its network service provider may not have direct physical connections to the Cloud DCs that are optimal for hosting the enterprise's specific workloads/Apps. Under those circumstances, SDWAN Overlay is a very flexible choice to interconnect the enterprise on-premises data centers & branch offices to its desired Cloud DCs.

However, SDWAN paths established over the public Internet can have unpredictable performance, especially over long distances and across operators' domains. Therefore, it is highly desirable to steer as much as possible the portion of SDWAN paths over the enterprise's existing VPN that has guaranteed SLA to minimize the distance or the number of segments over the public Internet.

MEF Cloud Service Architecture [MEF-Cloud] also describes a use case of network operators using SDWAN over LTE or the public Internet for last mile access where the VPN service providers cannot necessarily provide the required physical infrastructure.

Under those scenarios, one or two of the SDWAN endpoints may not be directly attached to the PEs of a VPN Domain.

When using SDWAN to connect the enterprise's existing sites to the workloads hosted in Cloud DCs w, the corresponding C-PEs have to be upgraded to support SDWAN. If the workloads hosted in Cloud DCs need to be connected to many sites, the upgrade process can be very expensive.

[Net2Cloud-Problem] describes a hybrid network approach that integrates SDWAN with traditional MPLS-based VPNs, to extend the existing MPLS-based VPNs to the Cloud DC Workloads over the access paths that are not under the VPN provider's control. To make it work properly, a small number of the PEs of the MPLS VPN can be designated to connect to the remote workloads via SDWAN secure IPsec tunnels. Those designated PEs are shown as fPE (floating PE or smart PE) in Figure 3. Once the secure IPsec tunnels are established, the workloads hosted in Cloud DCs can be reached by the enterprise's VPN without upgrading all of the enterprise's existing

CPEs. The only CPE that needs to support SDWAN would be a virtualized CPE instantiated within the cloud DC.

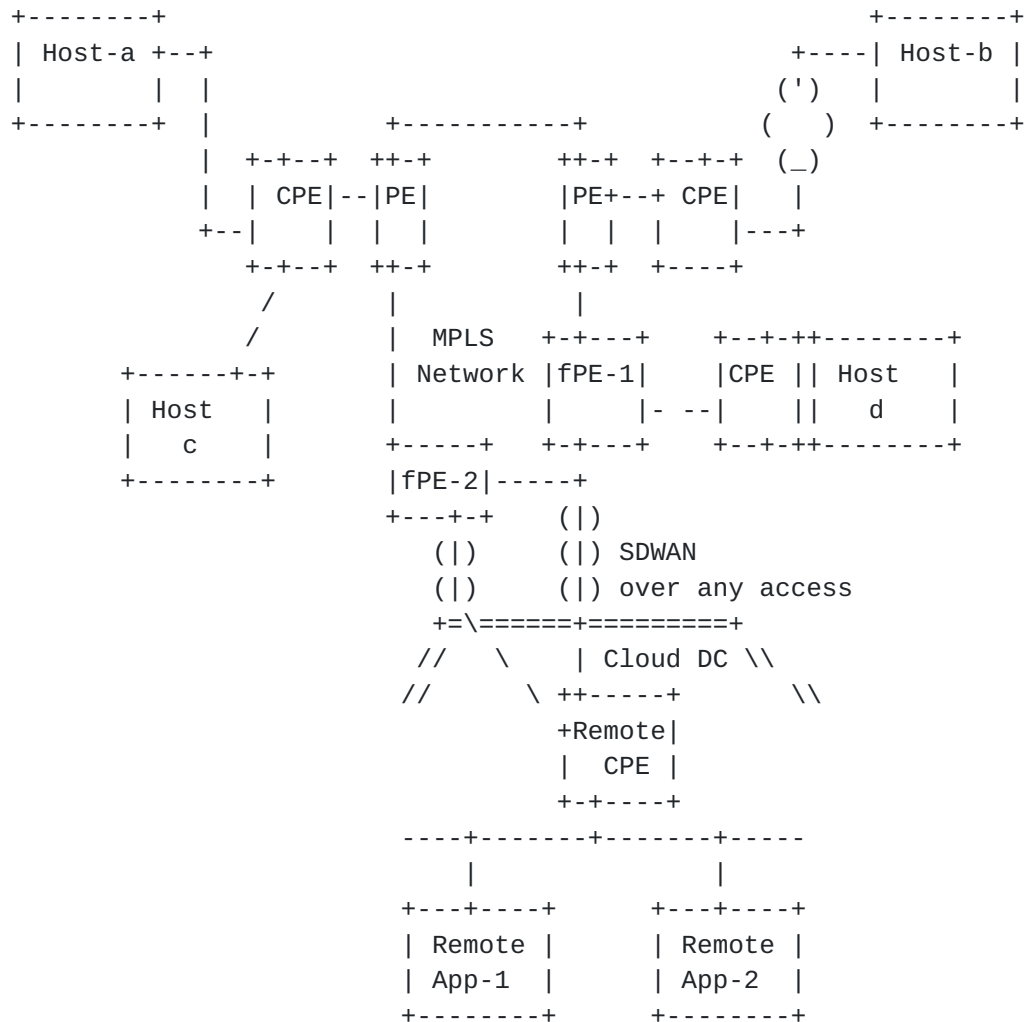


Figure 3: VPN Extension to Cloud DC

In Figure 3, the optimal Cloud DC to host the workloads (as a function of the proximity, capacity, pricing, or other criteria chosen by the enterprises) does not have a direct connection to the PEs of the MPLS VPN that interconnects the enterprise's existing sites.

5.1. Floating PEs to connect to Remote CPEs

To extend MPLS VPNs to remote CPEs, it is necessary to establish secure tunnels (such as IPsec tunnels) between the Floating PEs and the remote CPEs.

Even though a set of PEs can be manually selected to act as the floating PEs for a specific cloud data center, there are no standard protocols for those PEs to interact with the remote CPEs (most likely virtualized) instantiated in the third party cloud data centers (such as exchanging performance or route information).

When there is more than one fPE available for use (as there should be for resiliency purposes or the ability to support multiple cloud DCs geographically scattered), it is not straightforward to designate an egress fPE to remote CPEs based on applications. There is too much applications' traffic traversing PEs, and it is not feasible for PEs to recognize applications from the payload of packets.

5.2. NAT Traversal

Cloud DCs that only assign private IPv4 addresses to the instantiated workloads assume that traffic to/from the workload usually needs to traverse NATs.

A SDWAN edge node can solicit a 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 so that such information can be communicated to the relevant peers.

5.3. Complexity of using BGP between PEs and remote CPEs via Internet

Even though an EBGP (external BGP) Multi-hop design can be used to connect peers that are not directly connected to each other, there are still some complications in extending BGP from MPLS VPN PEs to remote CPEs via any access path (e.g., Internet).

The path between the remote CPEs and VPN PEs that maintain VPN routes may very well traverse untrusted nodes.

EBGP Multi-hop design requires static configuration on both peers. To use EBGP between a PE and remote CPEs, the PE has to be manually configured with the "next-hop" set to the IP address of the CPEs. When remote CPEs, especially remote virtualized CPEs are dynamically instantiated or removed, the configuration of Multi-Hop EBGP on the PE has to be changed accordingly.

Egress peering engineering (EPE) is not sufficient. Running BGP on virtualized CPEs in Cloud DCs requires GRE tunnels to be established first, which requires the remote CPEs to support address and key management capabilities. [RFC 7024](#) (Virtual Hub & Spoke) and Hierarchical VPN do not support the required properties.

Also, there is a need for a mechanism to automatically trigger configuration changes on PEs when remote CPEs' are instantiated or moved (leading to an IP address change) or deleted.

EBGP Multi-hop design does not include a security mechanism by default. The PE and remote CPEs need secure communication channels when connecting via the public Internet.

Remote CPEs, if instantiated in Cloud DCs, might have to traverse NATs to reach PEs. It is not clear how BGP can be used between devices located beyond the NAT and the devices located behind the NAT. It is not clear how to configure the Next Hop on the PEs to reach private IPv4 addresses.

[5.4.](#) Designated Forwarder to the remote edges

Among the multiple floating PEs that are reachable from a remote CPE, multicast traffic sent by the remote CPE towards the MPLS VPN can be forwarded back to the remote CPE due to the PE receiving the multicast packets forwarding the multicast/broadcast frame to other PEs that in turn send to all attached CPEs. This process may cause traffic loops.

Therefore, it is necessary to designate one floating PE as the CPE's Designated Forwarder, similar to TRILL's Appointed Forwarders [[RFC6325](#)].

MPLS VPNs do not have features like TRILL's Appointed Forwarders.

5.5. Traffic Path Management

When there are multiple floating PEs that have established IPsec tunnels with the remote CPE, the remote CPE can forward outbound traffic to the Designated Forwarder PE, which in turn forwards traffic to egress PEs and then to the final destinations. However, it is not straightforward for the egress PE to send back the return traffic to the Designated Forwarder PE.

Example of Return Path management using Figure 3 above.

- fPE-1 is DF for communication between App-1 <-> Host-a due to latency, pricing or other criteria.
- fPE-2 is DF for communication between App-1 <-> Host-b.

6. Manageability Considerations

Zero touch provisioning of SDWAN edge nodes should be a major feature of SDWAN deployments. It is necessary for a newly powered up SDWAN edge node to establish a secure connection (by means of TLS, DTLS, etc.) with its controller.

7. Security Considerations

The intention of this draft is to identify the gaps in current and proposed SDWAN approaches that can address requirements identified in [Net2Cloud-problem].

Several of these approaches have gaps in meeting enterprise security requirements when tunneling their traffic over the Internet, since this is the purpose of SDWAN. See the individual sections above for further discussion of these security gaps.

8. IANA Considerations

This document requires no IANA actions. RFC Editor: Please remove this section before publication.

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.

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, et al, "Subsequent Address Family Indicator for SDWAN Ports", [draft-dunbar-idr-sdwan-port-safi-00](#), Work-in-progress, March 2019.

[BGP-SDWAN-Usage] L. Dunbar, et al, "Framework of Using BGP for SDWAN Overlay Networks", [draft-dunbar-idr-sdwan-framework-00](#), work-in-progress, Feb 2019.

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

[SECURE-EVPN] A. Sajassi, et al, [draft-sajassi-bess-secure-evpn-01](#), work in progress, March 2019.

[SECURE-L3VPN] E. Rosen, "Provide Secure Layer L3VPNs over Public Infrastructure", [draft-rosen-bess-secure-l3vpn-00](#), work-in-progress, July 2018

[DMVPN] Dynamic Multi-point VPN:

<https://www.cisco.com/c/en/us/products/security/dynamic-multipoint-vpn-dmvpn/index.html>

[DSVPN] Dynamic Smart VPN:

<http://forum.huawei.com/enterprise/en/thread-390771-1-1.html>

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

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

10. Acknowledgments

Acknowledgements to John Drake for his review and contributions. Many thanks to John Scudder for stimulating the clarification discussion on the Tunnel-Encap draft so that our gap analysis can be more accurate.

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

Authors' Addresses

Linda Dunbar
Futurewei
Email: ldunbar@futurewei.com

Andrew G. Malis
Independent
Email: agmalis@gmail.com

Christian Jacquenet
Orange
Rennes, 35000
France
Email: Christian.jacquenet@orange.com