RTG Working Group
Internet Draft
Intended status: Informational
Expires: March 15, 2026

L. Dunbar
Futurewei
A. Malis
Malis Consulting
C. Jacquenet
Orange
M. Toy
Verizon
K. Majumdar
Oracle
September 15, 2025

Networks to Cloud DCs: Challenges and Mitigation Practices
draft-ietf-rtgwg-net2cloud-problem-statement-43

Abstract

This document describes a set of network-related problems enterprises face when interconnecting their branch offices with dynamic workloads in third-party data centers (DCs) (a.k.a. Cloud DCs). These challenges are particularly relevant to enterprises with conventional VPN services that want to leverage those networks (instead of altogether abandoning them).
The document also outlines various mitigation approaches, including those already developed within the IETF. For challenges that do not yet have established solutions, it identifies the IETF drafts that have been proposed to address these issues. The intent is to provide a cohesive view of problems and solution approaches that have been documented or proposed within the IETF.

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 March 15, 2026.

Copyright Notice

Table of Contents

1. Introduction

Cloud data centers (DCs) provide scalable, on-demand services across various geographic locations, enabling enterprises to deploy applications and workloads closer to users for improved latency. The dynamic nature of cloud workloads necessitates flexible networking solutions to accommodate changes in service locations and connectivity demands.

Cloud operators offer network functions such as virtual firewalls, private cloud services, and virtual PBX systems. As a shared infrastructure hosting multiple customers, Cloud DCs require enterprises to establish robust connectivity solutions to integrate existing VPNs with cloud networks.

This document examines networking challenges enterprises face when connecting branch offices to Cloud DCs and explores mitigation practices. While it references work from other standards development organizations (SDOs), its primary focus remains within the IETF's scope. Specifically, the document focuses on routing-related challenges that have active IETF discussions and proposed solution drafts, rather than attempting to address the entire problem space of enterprise-cloud connectivity. Individual IETF solution drafts address specific aspects of enterprise-cloud connectivity, but this document unifies these elements to provide a comprehensive perspective and emphasize the need for coordinated solutions. References to IETF working groups and Internet Drafts are included as examples to inform readers, without mandating the adoption of any specific solution. Section 6 outlines high-level, solution-agnostic requirements to guide future considerations in addressing these challenges.

2. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT","SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Terms used in this document are described below.

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

Hybrid Cloud: applications and workloads split among Cloud DCs owned or managed by different operators.

Hybrid Clouds: A hybrid cloud is a mixed computing environment where applications are run using a combination of computing, storage, and services in different public clouds and private clouds, including on-premises data centers or "edge" locations [HYBRID-CLOUD].

IXPs: Internet exchange points (IXes or IXPs) are the common grounds of IP networking, allowing participating Internet service providers (ISPs) to exchange data destined for their respective networks [WIKI-IXP].

Private Cloud: The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises. (NIST Special Publication 800-145).

SD-WAN An overlay connectivity service that optimizes transport of IP Packets over one or more Underlay Connectivity Services by recognizing applications (Application Flows) and determining forwarding behavior by applying Policies to them. [MEF-70.2]

VPC: A Virtual Private Cloud (VPC) is a secure, isolated segment of a public cloud, where users can deploy and manage resources such as virtual machines, databases, and applications. VPCs offer the flexibility of using the public cloud’s infrastructure while providing more control over networking and security.

3. Issues and Mitigation Methods of Connecting to Cloud DCs

This section identifies some high-level problems that can be addressed using IETF technologies and ongoing standardization efforts within the Routing area. Other Cloud DC related challenges, such as managing Cloud spending or issues outside the Routing scope, are out of the scope for this document.

3.1. Increased BGP Peering Errors and Mitigation Methods

Where conventional ISPs peer primarily with other ISPs and with a limited number of VPN enterprise customers, public Cloud DCs establish BGP sessions with a much larger and more diverse set of enterprise customers. Many of these enterprises and application providers are not experienced in managing complex BGP relationships, which increases the likelihood of configuration errors, such as capability mismatch, route leaks, missing Keepalives, and session resets. Capability mismatch, in particular, can cause BGP sessions not being adequately established. These issues are more acute for Cloud DCs than they have been, though they also affect conventional ISPs to a lesser degree.

BGP route convergence delays and security vulnerabilities, such as BGP hijacking, remain significant concerns when connecting enterprise networks to Cloud DCs. Route propagation policies and peering configurations vary across cloud providers, requiring enterprises to carefully design BGP session parameters to prevent route leaks, session resets, and excessive route advertisements. The use of BGP Route Reflectors, policy-based route filtering, and automated session monitoring can help mitigate these risks and improve BGP session stability in hybrid and multi-cloud environments.
Here are the recommended mitigation practices:

- A Cloud GW typically establishes multiple eBGP sessions with many clients. Each session is configured with a maximum number of routes it can handle. To avoid exceeding this limit, which could lead to the Cloud GW dropping routes, on-premises data center gateways should simplify their route advertisements by filtering unnecessary routes and using a default route instead. This practice minimizes the volume of routing information exchanged between on-premises data centers and Cloud DCs, thereby preventing the unwanted dropping of routes when the configured maximum for a client is exceeded, where appropriate and when consistent with enterprise policy and Cloud DC requirements.

- When a Cloud GW receives inbound routes exceeding the maximum routes from a peer, the current practice is to generate out-of-band alerts (e.g., Syslog entries) via the management system or to terminate the BGP session (with a cease notification message being sent per Section 4 of [RFC4486]). However, a more operation-friendly approach would be for peers to reduce the number of routes they are advertising. Therefore, it is worth considering adding a "route threshold crossing" alert mechanism to request peers to take action to reduce their advertised routes, rather than their BGP sessions being terminated by Cloud GW. While this mechanism is not available today and is beyond the scope of this document, further discussion in the IETF Inter-Domain Routing (IDR) Working Group is needed. Such work could lead to the addition of new subcodes in RFC4486 Section 3 and corresponding descriptions in RFC4486 Section 4 to facilitate this more efficient approach.
- If a Cloud GW, a BGP speaker, receives from its BGP peer a capability that it does not itself support or recognize, it MUST ignore that capability, and the BGP session MUST NOT be terminated per [RFC5492]. While ignoring unknown capabilities prevents unnecessary session resets, cloud operators should still monitor capability mismatches through logging or management systems to avoid configuration ambiguities.
- When receiving a BGP UPDATE with a malformed attribute, the revised BGP error handling procedure in [RFC7606] should be followed instead of session resetting.
- When a Cloud DC doesn’t support multi-hop eBGP peering with external devices, enterprise GWs need to establish tunnels (e.g., IPsec) to the Cloud GWs to form an IP neighbor relationship.

- Leveraging YANG models to programmatically synchronize configurations between BGP peers (e.g., [SVC-AC]) and to adjust the local configuration accordingly (e.g., [NTW-AC] or [DATAMODEL-BGP]). This proactive approach reduces the likelihood of BGP configuration issues and ensures that both BGP peers operate with synchronized and compatible settings, where YANG interfaces are supported.

3.2. Site Failures and Methods to Minimize Impacts

In this document, a site refers to a subdivision within a Cloud Data Center (Cloud DC), such as a building, a floor, a pod, or a server rack.

Failures within a site can include capacity degradation or complete out-of-service failure. Some examples of events that can trigger a site failure are: a) fiber cut for links connecting to the site or among pods within the site; b) cooling failures; c) insufficient backup power during a power failure; d) cyber threat attacks; e) too many changes outside of the maintenance window; etc. A fiber-cut is not uncommon in a Cloud DC or between DCs.

As described in [RFC7938], a Cloud DC may not run IGP within its domain, instead, it relies on internal methods to detect and report faults, which differ from standardized protocols like BFD or IGP. In the event of a site failure, while Cloud GW visible to clients continues to operate normally, the failure remains undetected by clients relying on BFD [RFC5880]. When BFD is not running within the Cloud DC, the GW cannot simply extend or concatenate BFD sessions to external peers.

When a site failure occurs, many services can be impacted. When the impacted services’ IP prefixes in a site are not well aggregated, which is common, one single site failure can trigger multiple BGP UPDATE messages. There are proposals, such as [METADATA-PATH], to enhance BGP advertisements to reduce the number of messages required.

[RFC7432] specifies a mass withdrawal mechanism for EVPN to signal a large number of routes being changed to remote PE nodes as quickly as possible. However, this alone is insufficient, as the routes at the sites might not all be EVPN routes.

3.3. Limitations of DNS-based Cloud DC Location Selection

Many applications have multiple instances running in different Cloud DCs. A commonly deployed solution has DNS server(s) responding to a Fully Qualified Domain Name (FQDN) inquiry with IP addresses of the instance in the closest or lowest cost DC. Here are some problems associated with DNS-based solutions:

- Dependent on client behavior
    - A misbehaving client can cache results indefinitely, even if the DNS TTL has expired.
    - Clients may fail to access a service even though there are servers available in other Cloud DCs because the failing IP address is still cached in the DNS resolver.
- No inherent awareness of proximity in the network (routing) layer, resulting in suboptimal performance.
- Inflexible traffic control: The Local DNS resolver becomes the unit of traffic management which requires DNS to receive periodic updates of the network condition, which can be operationally difficult.

One method to mitigate the problems listed above is to use anycast [RFC4786] for the services so that network proximity and conditions can be automatically considered in optimal path selection. However, anycast optimizes based on routing reachability and may not reflect real-time congestion or service load.

[METADATA-PATH] identifies metrics that can be utilized for the ingress routers to make path steering decisions not only based on the routing cost but also the running environment of the edge services. This complements DNS-based approaches by shifting decision-making to the routing layer.

[RFC8490] and [RFC8765] on stateful DNS can also help improve performance by refreshing the cache and handling session idle timeouts more effectively.

3.4. Network Issues for 5G Edge Clouds and Mitigation Methods

5G Edge Cloud DCs [3GPP-5G-Edge] may host edge computing applications for ultra-low latency services on virtual or physical servers. Those applications have low latency connections to the UEs

(User Equipment) and might have other connections to backend servers or databases in other locations.

The low latency traffic to/from the UEs is transported through the 5G gNB (Next Generation Node B), UPFs (User Plane Function) and the 5G Local Data Networks (LDN) to the edge Cloud DCs. The LDN’s ingress routers connected to the UPFs might be co-located with 5G Core functions in the edge Clouds. The 5G Core functions include Session Management Functions (SMF), Access Mobility Functions (AMF), User Plane Functions (UPF), and others.

Here are some network problems with connecting to the services in the 5G Edge Clouds:

1) While distances from the LDN Ingress router to server instances in different edge clouds may vary slightly, the overall service latency is significantly influenced by both routing distance and capacity status at the edge cloud. Therefore, a routing protocol solely based on the shortest routing distance alone may not guarantee the lowest overall latency. A more comprehensive approach that considers both factors is essential for the routing protocol to achieve service performance.
2) Due to user mobility, sources (UEs) can ingress from different LDN Ingress routers, presenting a routing challenge.

[METADATA-PATH] extends the BGP UPDATE messages for a Cloud GW to propagate the edge service-related metrics from Cloud GW to the ingress routers so that the ingress routers can incorporate the destination site’s capabilities with the routing distance in computing the optimal paths.

The IETF CATS (Computing-Aware Traffic Steering) working group is examining general aspects of this space and may come up with protocol recommendations for this information exchange.

3.5. DNS Practices for Hybrid Workloads

DNS name resolution is essential for on-premises and cloud-based resources. For customers with hybrid workloads, which include on-premises and cloud-based resources, extra steps are necessary to configure DNS to work seamlessly across both environments.

Each cloud operator has its own DNS to resolve resources within its Cloud DCs and to well-known public domains. A cloud DNS service can be configured to forward queries to customer managed authoritative DNS servers hosted on-premises and to respond to DNS queries forwarded by on-premises DNS servers.

For enterprises using multiple cloud providers, it is necessary to establish policies and rules on how/where to forward DNS queries. When applications in one Cloud need to communicate with applications hosted in another Cloud, DNS queries from one Cloud DC could be forwarded to the enterprises’ on-premises DNS, which in turn can be forwarded to the DNS service in another Cloud. Configuration can be complex depending on the application communication patterns.

However, name collisions can still occur even with carefully managed policies and configurations. Some organizations use internal names like those under a .internal top-level domain. However, .internal is not an officially designated special-use domain name by IANA nor an ICANN-approved Top-Level Domain. To avoid conflicts, enterprises should use a globally unique, registered domain name, even for internal resolution purposes. A globally unique name does not have to be globally resolvable. An organization’s domain can include subdomains that are only resolvable within restricted zones, zones that resolve differently depending on query origin, or zones that resolve consistently for all queries [Split-Horizon-DNS].

Using globally unique names prevents collisions and simplifies DNSSEC trust management, since registered domains can be chained to the global DNSSEC trust anchor. Enterprises should therefore consider using a registered FQDN from global DNS as the root for both enterprise and internal namespaces.

3.6. NAT Practices for Accessing Cloud Services

Cloud resources, such as VMs (Virtual Machine) or application instances, are commonly assigned with private IP addresses. When integrating multiple cloud environments or hybrid cloud architectures, enterprises often face overlapping private IP address spaces, requiring address translation techniques such as NAT. Managing NAT policies across different cloud providers can introduce additional complexity, particularly when ensuring consistent routing and avoiding conflicts between overlapping RFC1918 address ranges.

By configuration, some private subnets can have NAT functionality to reach out to external networks, while some private subnets are internal to a Cloud DC only.

Different cloud operators support different levels of NAT functionality. For example, in some environments a NAT gateway may not support connections through private endpoints, VPN, direct connections, or peering links [AWS-NAT]. In others, NAT services may provide outbound connectivity to the Internet for instances without public IP addresses, but not inbound NAT [Google-NAT]. These variations mean that enterprises must carefully evaluate provider-specific NAT features and limitations.

In addition to feature gaps across providers, NAT itself introduces operational challenges. Address translation can obscure end-to-end visibility, complicate troubleshooting, and make consistent policy enforcement more difficult across multiple domains. NAT state exhaustion and asymmetric routing can also lead to subtle service disruptions.

For enterprises with applications running in different Cloud DCs, NAT configurations must therefore be carefully coordinated across Cloud DCs and on-premises DCs to ensure consistency, prevent conflicts, and minimize operational complexity.

3.7. Cloud Discovery Practices

One of the concerns of enterprises using Cloud services is the lack of awareness of the locations of their services hosted in the Cloud, as cloud operators can move the service instances from one place to another. While geographic locations are usually exposed to enterprises, such as Availability Zones or Regions, the topological location is usually hidden. When applications in Cloud DCs communicate with on-premises applications, it may not be clear where the cloud applications are located or to which VPCs they belong.

Being able to detect cloud services’ location can help on-premises gateways (routers) to connect to services in a more optimal site, particularly when the enterprise’s end users or policies change.

For enterprises that instantiate virtual routers in Cloud DCs, metadata can be attached (e.g., GENEVE [RFC8926] header or IPv6 Extension Header) to indicate additional properties, including useful information about the sites where they are instantiated.

4. Dynamic Connecting Enterprise Sites with Cloud DCs

For many enterprises with established private VPNs (e.g., private circuits, MPLS-based L2VPN[RFC6136]/L3VPN[RFC4364]) interconnecting branch offices and on-premises data centers, connecting to Cloud services will be a mix of different types of networks. When an enterprise’s existing VPN service providers do not have direct connections to the desired Cloud DCs that the enterprise prefers to use, the enterprise faces additional infrastructure and operational costs to utilize the Cloud services.

This section describes some mechanisms for enterprises with private VPNs to connect to Cloud services dynamically.

4.1. Sites to Cloud DC

Most Cloud operators offer multiple types of network gateways (GWs) through which an enterprise can reach their workloads hosted in the Cloud DCs:

- Internet GW for services hosted in the Cloud DCs to be accessed by external requests via Internet routable addresses. E.g., AWS Internet GW [AWS-Cloud-WAN].
- IPsec tunnels terminating GW for establishing IPsec SAs [RFC6071] with an enterprise’s own gateway, so that the communications between those gateways can be secured from the underlay (which might be the public Internet). E.g., AWS Virtual gateway (vGW).
- Direct connect GW for enterprises to connect with Cloud services via private leased lines provided by Network Service Providers. E.g., AWS Direct Connect. In addition, an AWS Transit Gateway can be used to interconnect multiple VPCs in different Availability Zones. AWS Transit Gateway acts as a hub that controls how traffic is forwarded among all the connected networks which act like spokes.

Each cloud provider enforces its own routing mechanisms, such as AWS Transit Gateway, Azure Virtual WAN, and Google Cloud Dedicated Interconnect. These vendor-specific architectures create additional challenges for enterprises that require consistent routing policies across multiple cloud environments.

Microsoft Azure’s Virtual WAN [Azure-SD-WAN] allows extension of a private network to any of the Microsoft Cloud services, including Azure and Office365. ExpressRoute is configured using Layer 3 routing. Customers can opt for redundancy by provisioning dual links from their location to two Microsoft Enterprise edge routers (MSEEs) located within a third-party ExpressRoute peering location. The BGP routing protocol is then setup over WAN links to provide redundancy to the cloud. This redundancy is maintained from the peering data center into Microsoft's cloud network.

Google's Cloud Dedicated Interconnect offers similar network connectivity options as AWS and Microsoft. One distinct difference, however, is that Google's service allows customers access to the entire global Cloud network by default. It does this by connecting the on-premises network with the Google Cloud using BGP and Google Cloud Routers to provide optimal paths to the different regions of the global cloud infrastructure.

Figure 1 below shows an example of a portion of workloads belonging to one tenant (e.g., TN-1) that are accessible via a virtual router connected by AWS Internet Gateway; some of the same tenant (TN-1) services are accessible via AWS vGW, and others are accessible via AWS Direct Connect. The workloads belonging to one tenant can communicate within a Cloud DC via virtual routers (e.g., vR1, vR2).

Different types of access require different level of security functions. Sometimes it is not visible to end customers which type of network access is used for a specific application instance. To get better visibility, separate virtual routers (e.g., vR1 & vR2) can be deployed to differentiate traffic to/from different Cloud GWs. It is important for some enterprises to be able to observe the specific behaviors when connected by different connections.

A CPE (Customer Premises Equipment) can be a customer owned router or ports physically connected to an AWS Direct Connect GW.

```
     +------------------------+
     |    ,---.         ,---. |
     |   (TN-1 )       ( TN-2)|
     |    '-+-'  +---+  '-+-' |
     |      +----|vR1|----+   |
     |           ++--+        |
     |            |         +-+----+
     |            |        /Internet\ For external customers
     |            +-------+ Gateway  +----------------------
     |                     \        / to reach via Internet
     |                      +-+----+
     |                        |
     |    ,---.         ,---. |
     |   (TN-1 )       ( TN-2)|
     |    '-+-'  +---+  '-+-' |
     |      +----|vR2|----+   |
     |           ++--+        |
     |            |         +-+----+
     |            |        / virtual\ For IPsec Tunnel
     |            +-------+ Gateway  +----------------------
     |            |        \        /  termination
     |            |         +-+----+
     |            |           |
     |            |    + - - - - - - - - - - - - - - - --+
     |            |    |    +-+----+         +----+
     |            |        /        \ Direct /      \   |
     |            +----|--+ Gateway  +------+ Fabric|--VPN-- CPE
     |                     \        / Connect\ edge /   |
     |                 |    +-+----+         +----+
     |                        |         IXP             |
     |                 + - - - - - - - - - - - - - - - --+
     +------------------------+
```

     TN: Tenant Network. One TN can be attached to both vR1 and vR2.
     Figure 1: Examples of Multiple Cloud DC connections.

4.2. Inter-Cloud Connection

The connectivity options to Cloud DCs described in Section 4.1 are for reaching Cloud providers' DCs, but not between Cloud DCs. Inter-cloud routing complexity arises from the lack of standardized mechanisms for routing across multiple cloud providers. Each cloud operator applies distinct routing policies, which can create interoperability issues when establishing direct inter-cloud connections. Enterprises may leverage third-party cloud service brokers, SD-WAN overlays, or virtual routers instantiated in different Cloud DCs to optimize traffic flow across cloud environments.

Optimizing east-west traffic within and across Cloud DCs is critical for modern workloads, particularly for applications with high inter-service communication. Enterprises often rely on direct inter-VPC peering, SD-WAN overlays, or cloud-native transit services (e.g., AWS Transit Gateway, Azure Virtual WAN) to improve performance and reduce latency in multi-cloud and hybrid environments.

For example, when applications in AWS Cloud need to communicate with applications in Azure, today's practice requires a third-party gateway (physical or virtual) to interconnect the AWS's Layer 2 DirectConnect path with Azure's Layer 3 ExpressRoute.

Enterprises can also instantiate their virtual routers in different Cloud DCs and administer IPsec tunnels among them. In summary, here are some approaches, available to interconnect workloads among different Cloud DCs:

a) Utilize Cloud DC provided inter/intra-cloud connectivity services (e.g., AWS Transit Gateway) to connect workloads instantiated in multiple VPCs. Such services are provided with the Cloud gateway to connect to external networks (e.g., AWS DirectConnect Gateway).
b) Hairpin all traffic through the customer gateway, meaning all workloads are directly connected to the customer gateway, so that communications among workloads within one Cloud DC must traverse the customer gateway.
c) Establish direct tunnels among different VPCs (AWS' Virtual Private Clouds) and VNET (Azure's Virtual Networks) via client's own virtual routers instantiated within Cloud DCs. NHRP (Next Hop Resolution Protocol) [RFC2735] based multi-point techniques can be used to establish direct multi-point-to-Point or multi-point-to multi-point tunnels among those client's own virtual routers.
d) Utilize a Cloud Aggregator or Cloud Services Broker (CSB) who acts as an intermediary among cloud service providers and network service providers to offer a combined total package for enterprises. The Cloud Aggregator can provide the network connections among one enterprise's services instantiated in multiple Clouds.

Approach a) usually does not work if Cloud DCs are owned and managed by different Cloud providers.

Approach b) creates additional transmission delay plus incurring costs when exiting Cloud DCs.

For Approach c), [SDWAN-EDGE-DISCOVERY] describes a mechanism for virtual routers to advertise their properties for establishing proper IPsec tunnels among them. There could be other approaches developed to address the problem.

Approach d) is a method of third-party multi-cloud management business model.

4.3. Extending Private VPNs to Hybrid Cloud DCs

Traditional private VPNs, including private circuits or MPLS-based L2/L3 VPNs, have been widely deployed as an effective way to support businesses and organizations that require network performance and reliability although such services may be considered premium, available only at additional cost. Connecting an enterprise’s on-premise CPEs to a Cloud DC via a private VPN requires the private VPN provider to have a direct path to the Cloud GW. When the user base changes, the enterprise might want to migrate its workloads/applications to a new Cloud DC location closer to the new user base. The existing private VPN provider might not have circuits at the new location. Deploying PEs routers at new locations takes a long time (weeks, if not months).

When the private VPN network can’t reach the desired Cloud DCs, IPsec tunnels can dynamically connect the private VPN’s PEs with the desired Cloud DCs GWs. As the private VPNs provide higher quality of services, choosing a PE closest to the Cloud GW for the IPsec tunnel is desirable to minimize the IPsec tunnel distance over the public Internet.

In order to support Explicit Congestion Notification (ECN) [RFC3168] usage by private VPN traffic, the PEs that establish the IPsec tunnels with the Cloud GW need to comply with the ECN behavior specified by [RFC6040].

An enterprise can connect to multiple Cloud DC locations and establish different BGP peering with Cloud GW routers at different locations. As multiple Cloud DCs are interconnected by the Cloud provider’s own internal network, its topology and routing policies are not transparent or even visible to the enterprise customer’s on-premises routers. One Cloud GW BGP session might advertise all of the prefixes of the enterprise’s VPC, regardless of which Cloud DC a given prefix resides, which can cause improper optimal path selection for on-premises routers.

Managing hybrid cloud routing is further complicated by differences in cloud provider routing architectures, making consistent policy enforcement challenging. Enterprises often need to integrate SD-WAN solutions or other overlay technologies to harmonize routing behaviors across multiple cloud platforms.

To get around this problem, virtual routers in Cloud DCs can be used to attach metadata (e.g., in the GENEVE header or IPv6 Extension Header) to indicate the Geo-location of the Cloud DC, the delay measurement, or other relevant data.

5. Methods to Scale IPsec Tunnels to Cloud DCs

As described in Section 4.3, IPsec tunnels can be used to dynamically establish connection between private VPN PEs with Cloud GWs. Enterprises can also instantiate virtual routers within Cloud DCs to connect to their on-premises devices via IPsec tunnels.

As described in [Int-tunnels], IPsec tunnels can introduce MTU problems. This document assumes that endpoints manage the appropriate MTU sizes, therefore, not requiring VPN PEs to perform fragmentation when encapsulating user payloads in the IPsec packets.

5.1. Scale IPsec Tunnels Management

IPsec tunnels are a very convenient solution for an enterprise with a small number of locations to reach a Cloud DC. However, for a medium-to-large enterprise with multiple sites and data centers to fully connect to multiple Cloud DCs, there are N*C*2 bi-directional IPsec SAs (tunnels) between Cloud DC gateways and all those sites, with N being the number of enterprise sites and C being the number of Cloud sites. Each of those IPsec Tunnels requires pair-wise periodic key refreshment. For a company with hundreds or thousands of locations, managing hundreds (or even thousands) of IPsec tunnels can be very processing intensive. That is why many Cloud operators only allow a limited number of (IPsec) tunnels and bandwidth to each customer.

A solution like group key management [RFC4535] has been used to scale the IPsec key management. The group key management protocol documented in [RFC4535] outlines the relevant security risks for any group key management system in Section 3 (Security Considerations). While this particular protocol isn't being suggested, the drawbacks and risks of group key management are still relevant.

[SDWAN-EDGE-DISCOVERY] leverages the peers communication polices on the SD-WAN controller and BGP Update messages to exchange IPsec Security Associations related parameters among peers without IKEv2 point-to-point signaling or any other direct peer-to-peer session establishment messages.

5.2. CPEs Interconnection Over the Public Internet

When enterprise CPEs are far away from each other, e.g., across country/continent boundaries, the performance of IPsec tunnels over the public Internet can be problematic and unpredictable. Even though there are many monitoring tools available to measure delay and various performance characteristics of the network, the measurement for paths over the Internet is passive and past measurements may not represent future performance.

[MULTI-SEG-SDWAN] outlines some approaches for leveraging the Cloud backbone to connect enterprise CPEs across diverse geographical areas, eliminating the need for the Cloud GW to decrypt and re-encrypt traffic from the CPEs. A thorough examination of the security implications associated with this proposed method is necessary. Alternative encapsulations, like SRH (Segment Routing Header) [RFC8754] or others, can be considered for interconnecting enterprise CPEs.

6. Requirements for Networks Connecting Cloud Data Centers

To address the issues identified in this document, network solutions for connecting enterprises with their dynamic workloads or applications in Cloud DCs should satisfy the following requirements:

- Should support scalable policy management for the traffic to and from the newly instantiated application instances at any Cloud DC location. The scalable policy management, even though out of the scope of this document, can include centralized policy repositories and API-driven automation.

- Should allow enterprises to take advantage of the current state-of-the-art private VPN technologies, including the conventional circuit-based, MPLS-based VPNs, or IPsec-based VPNs (or any combination thereof) that run over the public Internet.
- Should support scalable IPsec key management among all nodes involved in DC interconnect schemes.
- Should support easy and fast, on-demand network connections to dynamic workloads and applications in Cloud DCs and easily reach these workloads when they migrate within or across data centers.
- Should support traffic steering to distribute loads across regions or Availability Zones based on performance/availability of workloads in addition to the network path conditions to the Cloud DCs.
- Should support network traffic traceability, logging, and diagnostics.
- Should support transit/spoke gateways interconnection scalability and consistent policy enforcement as workloads are increased/migrated. This requirement is mainly for the Cloud Aggregators or Cloud Service Brokers who provide managed services to enterprises over multiple Cloud service providers.

7. Security Considerations

This document focuses on security challenges directly related to networking and routing in enterprise-cloud connectivity, rather than broader cloud security concerns such as encryption at rest, patch management, and regulatory compliance. While those aspects are important, they fall outside the scope of this document, which specifically highlights network security risks, including BGP security, DDoS mitigation, VPN scalability, and inter-cloud connectivity risks. The security issues in terms of networking to Cloud DCs include:

- Service instances in Cloud DCs are connected to users (enterprises) via Public IP ports which are exposed to the following security risks:

a) Potential DDoS (Distributed Denial of Service) attack to the ports facing the untrusted network (e.g., the public Internet), which may propagate to the cloud edge resources. To mitigate such security risk, it is necessary for the ports facing Internet to enable Anti-DDoS features. There are many Anti-DDoS features to consider. Some examples include Rate Limiting, Access Control Lists (ACLs), Deep Packet Inspection (DPI), Blackholing and Sinkholing (which route malicious traffic to a non-existent IP address or a system that safely absorbs or analyzes the traffic), Traffic Scrubbing, and Geo-IP Blocking.

b) Potential risk of augmenting the attack surface with inter-Cloud DC connection by means of identity spoofing, man-in-the-middle, eavesdropping or DDoS attacks. One example of mitigating such attacks is using DTLS to authenticate and encrypt MPLS-in-UDP encapsulation [RFC7510].

- Potential attacks from service instances within the cloud. For example, data breaches, compromised credentials, and broken authentication, hacked interfaces and APIs, and account hijacking.

- When IPsec tunnels established from enterprise on-premises CPEs are terminated at the Cloud DC gateway where the workloads or applications are hosted, traffic to/from an enterprise’s workload can be exposed to others behind the data center gateway (e.g., exposed to other organizations that have workloads in the same data center).

  To ensure that traffic to/from workloads is not exposed to unwanted entities, IPsec tunnels may go all the way to the workload (servers, or VMs) within the DC.

- BGP security risks, including BGP hijacking and route leaks, can lead to malicious traffic redirection. To mitigate these risks, enterprises should implement BGP authentication (e.g., TCP MD5 or GTSM), RPKI for route validation, and strict inbound/outbound route filtering. Additionally, session security measures, such as RFC5492 for handling unsupported BGP capabilities and RFC7606 for improved error handling, can enhance routing stability and resilience.

- Group key management [RFC4535] comes with security risks such as: keys being used too long, single points of compromise (one compromise affects the whole group), key distribution vulnerabilities, key generation vulnerabilities, to name a few.

  [RFC4535] outlines the security risks in Section 3 (Security Considerations). While [RFC4535] specific protocol isn't being suggested, the risks and vulnerabilities apply to any group key management system.

- Striking a balance between scaling IPsec tunnel management outlined in this document and maintaining robust security is a delicate consideration. Simplifying the IPsec tunnel management to reduce management complexity for large SD-WAN networks might come with the inherent risk of decreased security. Careful consideration of the specific deployments, coupled with regular security assessments, is crucial to ensure the integrity and confidentiality of the transmitted data.

The Cloud DC operator's security practices can affect the overall security posture and need to be evaluated by customers. Many Cloud operators offer monitoring services for data stored in Clouds, such as AWS CloudTrail, Azure Monitor, and many third-party monitoring tools to improve the visibility of data stored in Clouds.

Solution drafts resulting from this work will address security concerns inherent to the solution(s), including both protocol aspects and the importance, for example, of securing workloads in Cloud DCs and the use of secure interconnection mechanisms.

A full security evaluation will be needed before [MULTI-SEG-SDWAN] and [SDWAN-EDGE-DISCOVERY] can be recommended as a solution to some problems described in this document.

8. IANA Considerations

This document requires no IANA actions.

9. References

9.1. Normative References

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

[RFC3168] K. Ramakrishnan, et al, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC3168, Sept. 2001.

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

[RFC4486] E. Chen and V. Gillet, "Subcodes for BGP Cease Notification Message", RFC4486, April 2006.

[RFC4535] H. Harney, et a, "GSAKMP: Group Secure Association Key Management Protocol", RFC4535, June 2006.

[RFC4786] J. Abley and K. Lindqvist, "Operation of Anycast Services", RFC4786, Dec. 2006.

[RFC5492] J. Scudder and R. Chandra, "Capabilities Advertisement with BGP-4", RFC5492, Feb. 2009.

[RFC5880] D. Katz and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC5880, June 2010.

[RFC6040] B. Briscoe, "Tunnelling of Explicit Congestion Notification", RFC6040, Nov 2010.

[RFC6136] A. Sajassi and D. Mohan, "Layer 2 Virtual Private Network (L2VPN) Operations, Administration, and Maintenance (OAM) Requirements and Framework", RFC6136, March 2011.

[RFC7606] E. Chen, et al "Revised Error Handling for BGP UPDATE Messages". Aug 2015.

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

[RFC7510] X. Xu, et al, "Encapsulating MPLS in UDP", RFC7510, April, 2015.

[RFC7938] P. Lapukhov, "Use of BGP for Routing in Large-Scale Data Centers", RFC7938, Aug. 2016.

[RFC8174] B. Leiba, "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", RFC8174, May 2017.

[RFC8490] R. Bellis, et al, "DNS Stateful Operations", RFC8490, March 2019.

[RFC8754] C. Filsfils, et al, "IPv6 Segment Routing Header (SRH)", RFC8754, March 2020.

[RFC8765] T. Pusateri and S. Cheshire, "DNS Push Notifications", RFC8765, June 2020.

[RFC8926] J. Gross and T. Sridhar, "Geneve: Generic Network Virtualization Encapsulation", RFC8926, Nov. 2020.

9.2. Informative References

[RFC2735] B. Fox, et al "NHRP Support for Virtual Private networks". Dec. 1999.

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

[3GPP-5G-Edge] 3GPP TS 23.548 v18.1.1, "5G System Enhancements for Edge Computing", April 2023.

[SDWAN-EDGE-DISCOVERY] L. Dunbar, S. Hares, R. Raszuk, K. Majumdar, G. Mishra, V. Kasiviswanathan, "BGP UPDATE for SD-WAN Edge Discovery", draft-ietf-idr-sdwan-edge-discovery-20, Jan 2025.

[AWS-NAT] NAT gateways - Amazon Virtual Private Cloud.

[AWS-Cloud-WAN] Introducing AWS Cloud WAN (Preview) | Networking & Content Delivery (amazon.com).

[Azure-SD-WAN] Architecture: Virtual WAN and SD-WAN connectivity - Azure Virtual WAN | Microsoft Learn.

[NTW-AC] M. Boucadair, et al, "A Network YANG Data Model for Attachment Circuits", draft-ietf-opsawg-ntw-attachment-circuit-12, July 2024.

[DATAMODEL-BGP] M. Jethanandani, K. Patel, S. Hares, "YANG Model for Border Gateway Protocol (BGP-4)", draft-ietf-idr-bgp-model-17, July 2023.

[Google-NAT] https://cloud.google.com/nat/docs/overview

[HYBRID-CLOUD] https://cloud.google.com/learn/what-is-hybrid-cloud

[Int-tunnels] J. Touch and W Townsley, "IP Tunnels in the Internet Architecture", draft-ietf-intarea-tunnels-13.txt, March 2023.

[MEF-70.2] MEF 70.2 SD-WAN Service Attributes and Service Framework. Oct. 2023.

[METADATA-PATH] L. Dunbar, et al, "BGP Extension for 5G Edge Service Metadata" draft-ietf-idr-5g-edge-service-metadata-26, Jan, 2025.

[MULTI-SEG-SDWAN] K. Majumdar, et al, "Multi-segment SD-WAN via Cloud DCs", draft-ietf-rtgwg-multisegment-sdwan-01, June 2024.

[SVC-AC] M. Boucadair, et al. "YANG Data Models for 'Attachment Circuits'-as-a-Service (ACaaS)", draft-ietf-opsawg-teas-attachment-circuit-19, Jan 2025.

[Split-Horizon-DNS] K. Tirumaleswar, et al, "Establishing Local DNS Authority in Validated Split-Horizon Environments", draft-ietf-add-split-horizon-authority-14, June 2024.

[WIKI-IXP] https://en.wikipedia.org/wiki/Internet_exchange_point

10. Acknowledgments

Many thanks to Joel Halpern, Aseem Choudhary, Adrian Farrel, Alia Atlas, Chris Bowers, Mohamed Boucadair, Paul Vixie, Paul Ebersman, Timothy Morizot, Ignas Bagdonas, Donald Eastlake, Michael Huang, Liu Yuan Jiao, Katherine Zhao, and Jim Guichard for the discussion and contributions.

Authors' Addresses

Linda Dunbar
Futurewei
Email: Linda.Dunbar@futurewei.com

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

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

Mehmet Toy
Verizon
One Verizon Way
Basking Ridge, NJ 07920
Email: mehmet.toy@verizon.com

Kausik Majumdar
Microsoft Azure
kmajumdar@microsoft.com