Network Working Group   L. Dunbar
Internet Draft   A. Malis
Intended status: Informational   Huawei
Expires: July 2019   C. Jacquenet
                                                                 Orange
                                                                 M. Toy
                                                                Verizon
                                                      February 12, 2019

       Seamless Interconnect Underlay to Cloud Overlay Problem Statement
                draft-ietf-rtgwg-net2cloud-problem-statement-00

Abstract

   This document describes the problems that enterprises face today
   when connecting their branch offices to dynamic workloads in third
   party data centers (a.k.a. Cloud DCs).

   It examines some of the approaches interconnecting cloud DCs with
   enterprises' on-premises DCs & branch offices. This document also
   describes some of the (network) problems that many enterprises face
   when they have workloads & applications & data split among hybrid
   data centers, especially for those enterprises with multiple sites
   that are already interconnected by VPNs (e.g., MPLS L2VPN/L3VPN).

   Current operational problems are examined to determine whether there
   is a need to improve existing protocols or whether a new protocol is
   necessary to solve them.

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 August 6, 2019.

Copyright Notice

Table of Contents

1. Introduction

   The ever-increasing use of cloud applications for communication
   services change the way corporate business works and shares
   information. Such cloud applications use resources hosted in third
   party DCs that also host services for other customers.

   With the advent of widely available third party cloud DCs in diverse
   geographic locations and the advancement of tools for monitoring and
   predicting application behaviors, it is technically feasible for
   enterprises to instantiate applications and workloads in locations
   that are geographically closest to their end-users. Such proximity
   improves end-to-end latency and overall user experience. Conversely,
   an enterprise can easily shutdown applications and workloads
   whenever end-users are in motion (thereby modifying the networking
   connection of subsequently relocated applications and workloads). In
   addition, an enterprise may wish to take advantage of more and more
   business applications offered by third party private cloud DCs.

   Most of those enterprise branch offices & on-premises data centers
   are already connected via VPNs, such as MPLS-based L2VPNs and
   L3VPNs. Then connecting to the cloud-hosted resources may not be
   straightforward if the provider of the VPN service does not have
   direct connections to the corresponding cloud DCs. Under those
   circumstances, the enterprise can upgrade the CPEs deployed in its
   various premises to utilize SD-WAN techniques to reach cloud
   resources (without any assistance from the VPN service provider), or
   wait for their VPN service provider to make new agreements with data

center providers to connect to the cloud resources. Either way has
additional infrastructure and operational costs.

In addition, it is an uptrend with more enterprises instantiating
their apps & workloads in different cloud DCs to maximize the
benefits of geographical proximity, elasticity and special features
offered by different cloud DCs.

2. Definition of terms

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

Controller: Used interchangeably with SD-WAN controller to manage
             SD-WAN overlay path creation/deletion and monitoring the
             path conditions between two or more sites.

DSVPN: Dynamic Smart Virtual Private Network. DSVPN is a secure
             network that exchanges data between sites without
             needing to pass traffic through an organization's
             headquarter virtual private network (VPN) server or
             router.

Heterogeneous Cloud: applications & workloads split among Cloud DCs
             owned & managed by different operators.

Hybrid Clouds: Hybrid Clouds (usually plural) refer to enterprises
             using their own premises DCs in addition to Cloud
             services provided by multiple cloud operators. For
             example, an enterprise not only have applications
             running in their own DCs, but also have applications
             hosted in multiple third party cloud DCs ((AWS, Azure,
             Google, Salesforces, SAP, etc). . ONUG also has a
             notion of heterogeneous cloud, refers to enterprises
             does not have its own DC, only uses services by 3rd
             party cloud operators.

SD-WAN: Software Defined Wide Area Network. In this document,
             "SD-WAN" refers to the solutions specified by ONUG (Open
             Network User Group), https://www.onug.net/software-

defined-wide-area-network-sd-wan/, which is about
pooling WAN bandwidth from multiple underlay networks to
get better WAN bandwidth management, visibility &
control. When the underlay networks are private
networks, traffic can traverse without additional
encryption; when the underlay networks are public, such
as Internet, some traffic needs to be encrypted when
traversing through (depending on user provided
policies).

VPC: Virtual Private Cloud. A service offered by Cloud DC
operators to allocate logically-isolated cloud
resources, including compute, networking and storage.

3. Current Practices in Interconnecting Enterprise Sites with Cloud DCs

3.1. Interconnect to Cloud DCs

Most Cloud operators offer some type of network gateway through
which an enterprise can reach their workloads hosted in the Cloud
DCs. For example, AWS (Amazon Web Services) offers the following
options to reach workloads in AWS Cloud DCs:

- Internet gateway for any external entities to reach the
  workloads hosted in AWS Cloud DC via the Internet.
- Virtual gateway (vGW) where IPsec tunnels [RFC6071] are
  established between an enterprise's own gateway and AWS vGW, so
  that the communications between those gateways can be secured
  from the underlay (which might be the public Internet).
- Direct Connect, which allows enterprises to purchase direct
  connect from network service providers to get a private leased
  line interconnecting the enterprises gateway(s) and the AWS
  Direct Connect routers. Via Direct Connect, an AWS Transit
  Gateway can be used to interconnect multiple VPCs in different
  Availability Zones.

CPEs at one Enterprise branch office are connected to the Internet
to reach AWS's vGW via IPsec tunnels. Other ports of such CPEs are
connected to AWS DirectConnect via a private network (without any
encryption).

```
    +-----------------------+
    |     ,---.          ,---. |
    |   (TN-1 )        ( TN-2)|
    |     `-+-'   +--+   `-+-' |
    |      +----|vR|-----+   |
    |           ++-+         |
    |           |        +-+----+
    |           |        /Internet\ For External
    |        +-------+ Gateway  +---------------------
    |                 \         / to reach via Internet
    |                  +-+----+
    |                     |
    +-----------------------+


    +-----------------------+
    |     ,---.          ,---. |
    |   (TN-1 )        ( TN-2)|
    |     `-+-'   +--+   `-+-' |
    |      +----|vR|-----+   |
    |           ++-+         |
    |           |        +-+----+
    |           |        / virtual\ For IPsec Tunnel
    |        +-------+ Gateway  +---------------------
    |                 \         /  termination
    |                  +-+----+
    |                     |
    +-----------------------+


    +-----------------------+
    |     ,---.          ,---. |
    |   (TN-1 )        ( TN-2)|
    |     `-+-'   +--+   `-+-' |
    |      +----|vR|-----+   |
    |           ++-+         |
    |           |        +-+----+              +------+
    |           |        /         \ For Direct /customer\
    |        +-------+ Gateway  +----------+ gateway  |
    |                 \         / Connect   \        /
    |                  +-+----+              +------+
    |                     |
    +-----------------------+
```
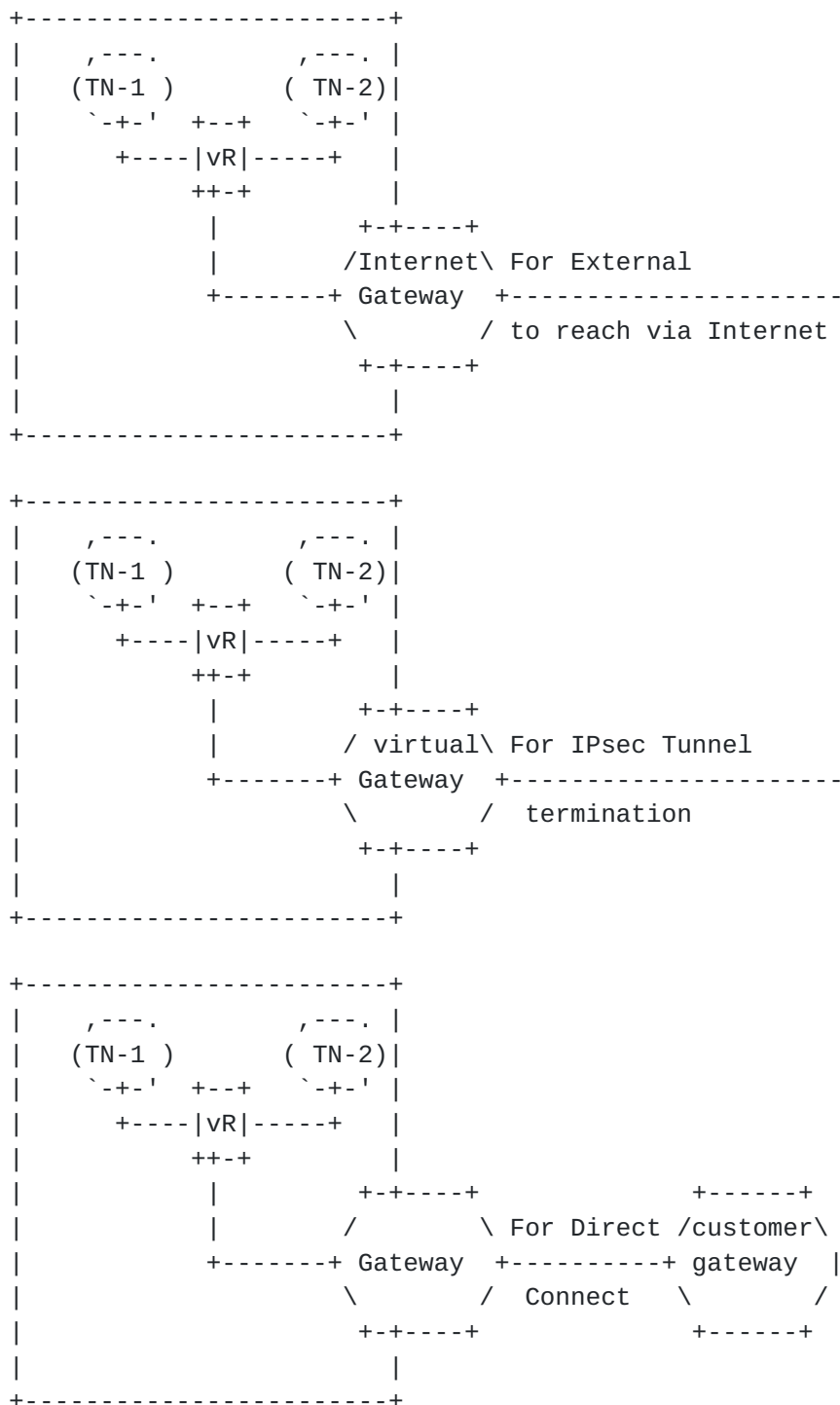
                Figure 1: Examples of Cloud DC connections.

3.2. Interconnect to Hybrid Cloud DCs

According to Gartner, by 2020 "hybrid will be the most common usage of the cloud" as more enterprises see the benefits of integrating public and private cloud infrastructures. However, enabling the growth of hybrid cloud deployments in the enterprise requires fast and safe interconnection between public and private cloud services. For an enterprise to connect to applications & workloads hosted in multiple Cloud DCs, the enterprise can use IPsec tunnels established over the Internet or a (virtualized) leased line service to connect its on-premises gateways to each of the Cloud DC's gateways, virtual routers instantiated in the Cloud DCs, or any other suitable design (including a combination thereof).

Some enterprises prefer to instantiate their own virtual CPEs/routers inside the Cloud DC to connect the workloads within the Cloud DC. Then an overlay path is established between customer gateways to the virtual CPEs/routers for reaching the workloads inside the cloud DC.

3.3. Connecting workloads among hybrid Cloud DCs

There are multiple approaches to interconnect workloads among different Cloud DCs:

  - Utilize Cloud DC provided transit gateways, which usually does not work if Cloud DCs are owned and managed by different Cloud providers.
  - Hairpin all the traffic through the customer gateway, which creates additional transmission delay & incurs cost when exiting Cloud DCs, or
  - Establish direct tunnels among different VPCs (Virtual Private Clouds) via client's own virtual routers instantiated within Cloud DCs. DMVPN (Dynamic Multipoint Virtual Private Network) or DSVPN (Dynamic Smart VPN) 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.

DMVPN & DSVPN use NHRP (Next Hop Resolution Protocol) [RFC2735] so that spoke nodes can register their IP addresses & WAN ports with the hub node. The IETF ION (Internetworking over NBMA (non-broadcast

multiple access) WG standardized NHRP for connection-oriented NBMA network (such as ATM) network address resolution more than two decades ago.

There are many differences between virtual routers in Public Cloud DCs and the nodes in an NBMA network. NHRP & DSVPN are not cannot be used for registering virtual routers in Cloud DCs unless an extension of such protocols is developed for that purpose. Other protocols such as BGP can be used, as described in [BGP-SDWAN].

[4]. **Desired Properties for Networks that interconnect Hybrid Clouds**
   The networks that interconnect hybrid cloud DCs must address the
   following requirements:
     - High availability at any time, whatever the duration of the
        connection to the cloud DC.
        Many enterprises include cloud infrastructures in their
        disaster recovery strategy, e.g., by enforcing periodic backup
        policies within the cloud, or by running backup applications in
        the Cloud, etc. Therefore, the connection to the cloud DCs may
        not be permanent, but rather needs to be on-demand.

     - Global reachability from different geographical zones, thereby
        facilitating the proximity of applications as a function of the
        end users' location, to improve latency.
     - Elasticity and mobility, to instantiate additional applications
        at Cloud DCs when end-users' usages increase and shut down
        applications at locations when there are fewer end-users.
        Some enterprises have front-end web portals running in cloud
        DCs and database servers in their on-premises DCs. Those Front-
        end web portals need to be reachable from the public Internet.
        The backend connection to the sensitive data in database
        servers hosted in the on-premises DCs might need secure
        connections.

     - Scalable security management. IPsec is commonly used to
        interconnect cloud gateways with CPEs deployed in the
        enterprise premises. For enterprises with a large number or
        branch offices, managing the IPsec's Security Associations
        among many nodes can be very difficult.

[5]. Problems with MPLS-based VPNs extending to Hybrid Cloud DCs

   Traditional MPLS-based VPNs have been widely deployed as an
   effective way to support businesses and organizations that require
   network performance and reliability. MPLS shifted the burden of
   managing a VPN service from enterprises to service providers. The
   CPEs attached to MPLS VPNs are also simpler and less expensive,
   since they do not need to manage routes to remote sites; they simply
   pass all outbound traffic to the MPLS VPN PEs to which the CPEs are
   attached (albeit multi-homing scenarios require more processing
   logic on CPEs). MPLS has addressed the problems of scale,
   availability, and fast recovery from network faults, and
   incorporated traffic-engineering capabilities.

   However, traditional MPLS-based VPN solutions are sub-optimized for
   connecting end-users to dynamic workloads/applications in cloud DCs
   because:

     - The Provider Edge (PE) nodes of the enterprise's VPNs might not
        have direct connections to third party cloud DCs that are used
        for hosting workloads with the goal of providing an easy access
        to enterprises' end-users.

     - It usually takes some time to deploy provider edge (PE) routers
        at new locations. When enterprise's workloads are changed from
        one cloud DC to another (i.e., removed from one DC and re-
        instantiated to another location when demand changes), the
        enterprise branch offices need to be connected to the new cloud
        DC, but the network service provider might not have PEs located
        at the new location.

        One of the main drivers for moving workloads into the cloud is
        the widely available cloud DCs at geographically diverse
        locations, where apps can be instantiated so that they can be
        as close to their end-users as possible. When the user base
        changes, the applications may be migrated to a new cloud DC
        location closest to the new user base.

- Most of the cloud DCs do not expose their internal networks, so
  the MPLS-based VPNs can only reach Cloud DC's Gateways, not to
  the workloads hosted inside.

- Many cloud DCs use an overlay to connect their gateways to the
  workloads located inside the DC. There has not been any
  standard to address the interworking between the Cloud Overlay
  and the enterprise' existing underlay networks.

Another roadblock is the lack of a standard way to express and
enforce consistent security policies for workloads that not only use
virtual addresses, but in which are also very likely hosted in
different locations within the Cloud DC [RFC8192]. The current VPN
path computation and bandwidth allocation schemes may not be
flexible enough to address the need for enterprises to rapidly
connect to dynamically instantiated (or removed) workloads and
applications regardless of their location/nature (i.e., third party
cloud DCs).

6. Problem with using IPsec tunnels to Cloud DCs

As described in the previous section, many Cloud operators expose
their gateways for external entities (which can be enterprises
themselves) to directly establish IPsec tunnels. Enterprises can
also instantiate virtual routers within Cloud DCs to connect to
their on-premises devices via IPsec tunnels. If there is only one
enterprise location that needs to reach the Cloud DC, an IPsec
tunnel is a very convenient solution.

However, many medium-to-large enterprises usually have multiple
sites and multiple data centers. For workloads and apps hosted in
cloud DCs, multiple sites need to communicate securely with those
cloud workloads and apps. This section documents some of the issues
associated with using IPsec tunnels to connect enterprise premises
with cloud gateways.

6.1. Complexity of multi-point any-to-any interconnection

The dynamic workload instantiated in cloud DC needs to communicate
with multiple branch offices and on-premises data centers. Most
enterprises need multi-point interconnection among multiple
locations, which can be provided by means of MPLS L2/L3 VPNs.

Using IPsec overlay paths to connect all branches & on-premises data centers to cloud DCs requires CPEs to manage routing among Cloud DCs gateways and the CPEs located at other branch locations, which can dramatically increase the complexity of the design, possibly at the cost of jeopardizing the CPE performance.

The complexity of requiring CPEs to maintain routing among other CPEs is one of the reasons why enterprises migrated from Frame Relay based services to MPLS-based VPN services.

MPLS-based VPNs have their PEs directly connected to the CPEs. Therefore, CPEs only need to forward all traffic to the directly attached PEs, which are therefore responsible for enforcing the routing policy within the corresponding VPNs. Even for multi-homed CPEs, the CPEs only need to forward traffic among the directly connected PEs. However, when using IPsec tunnels between CPEs and Cloud DCs, the CPEs need to compute, select, establish and maintain routes for traffic to be forwarded to Cloud DCs, to remote CPEs via VPN, or directly.

6.2. Poor performance over long distance

When enterprise CPEs or gateways are far away from cloud DC gateways or across country/continent boundaries, 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.

Many cloud providers can replicate workloads in different available zones. An App instantiated in a cloud DC closest to clients may have to cooperate with another App (or its mirror image) in another region or database server(s) in the on-premises DC. This kind of coordination requires predicable networking behavior/performance among those locations.

6.3. Scaling Issues with IPsec Tunnels

IPsec can achieve secure overlay connections between two locations over any underlay network, e.g., between CPEs and Cloud DC Gateways.

If there is only one enterprise location connected to the cloud gateway, a small number of IPsec tunnels can be configured on-demand

between the on-premises DC and the Cloud DC, which is an easy and
flexible solution.

However, for multiple enterprise locations to reach workloads hosted
in cloud DCs, the cloud DC gateway needs to maintain multiple IPsec
tunnels to all those locations (e.g., as a hub & spoke topology).
For a company with hundreds or thousands of locations, there could
be hundreds (or even thousands) of IPsec tunnels terminating at the
cloud DC gateway, which is not only very expensive (because Cloud
Operators usually charge their customers based on connections), but
can be very processing intensive for the gateway. Many cloud
operators only allow a limited number of (IPsec) tunnels & bandwidth
to each customer. Alternatively, you could use a solution like
group encryption where a single IPsec SA is necessary at the GW but
the drawback here is key distribution and maintenance of a key
server, etc.

[7](). Problems of Using SD-WAN to connect to Cloud DCs
   SD-WAN can establish parallel paths over multiple underlay networks
   between two locations on-demand, for example, to support the
   connections established between two CPEs interconnected by a
   traditional MPLS VPN ([RFC4364] or [RFC4664]) or by IPsec [RFC6071]
   tunnels.

   SD-WAN lets enterprises augment their current VPN network with cost-
   effective, readily available Broadband Internet connectivity,
   enabling some traffic offloading to paths over the Internet
   according to differentiated, possibly application-based traffic
   forwarding policies, or when the MPLS VPN connection between the two
   locations is congested, or otherwise undesirable or unavailable.

[7.1](). SD-WAN among branch offices vs. interconnect to Cloud DCs

   SD-WAN interconnection of branch offices is not as simple as it
   appears. For an enterprise with multiple sites, using SD-WAN overlay
   paths among sites requires each CPE to manage all the addresses that
   local hosts have the potential to reach, i.e., map internal VPN
   addresses to appropriate SD-WAN paths. This is similar to the
   complexity of Frame Relay based VPNs, where each CPE needed to
   maintain mesh routing for all destinations if they were to avoid an
   extra hop through a hub router. Even though SD-WAN CPEs can get
   assistance from a central controller (instead of running a routing
   protocol) to resolve the mapping between destinations and SD-WAN
   paths, SD-WAN CPEs are still responsible for routing table

maintenance as remote destinations change their attachments, e.g.,
the dynamic workload in other DCs are de-commissioned or added.

Even though originally envisioned for interconnecting branch
offices, SD-WAN offers a very attractive way for enterprises to
connect to Cloud DCs.

The SD-WAN for interconnecting branch offices and the SD-WAN for
interconnecting to Cloud DCs have some differences:

   - SD-WAN for interconnecting branch offices usually have two end-
     points (e.g., CPEs) controlled by one entity (e.g., a
     controller or management system operated by the enterprise).
   - SD-WAN for Cloud DC interconnects may consider CPEs owned or
     managed by the enterprise, while remote end-points are being
     managed or controlled by Cloud DCs (For the ease of
     description, let's call such CPEs asymmetrically-managed CPEs).

- Cloud DCs may have different entry points (or devices) with one
  entry point that terminates a private direct connection (based
  upon a leased line for example) and other entry points being
  devices terminating the IPsec tunnels, as shown in Figure 2.

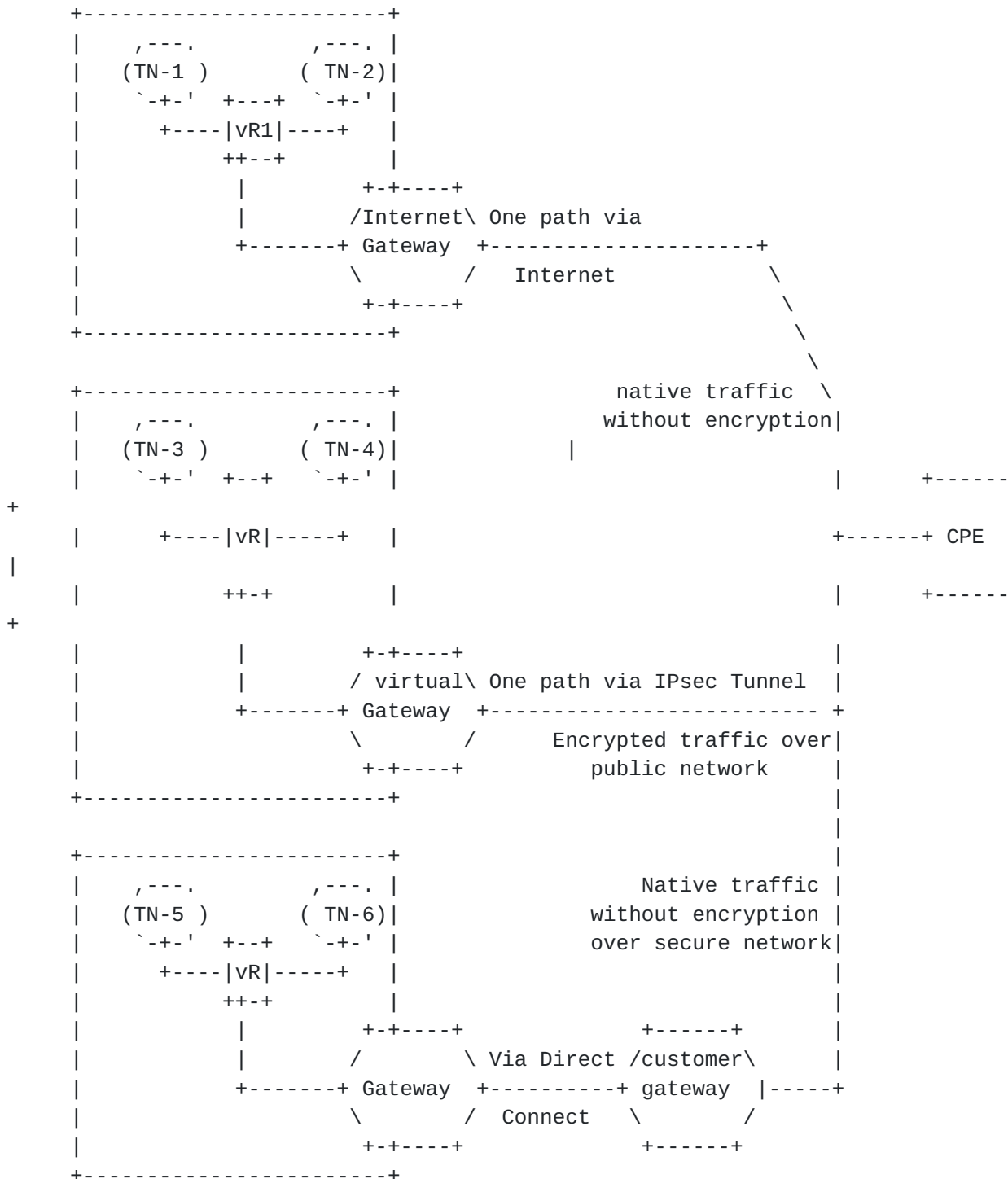Therefore, the SD-WAN design becomes asymmetric.

```
+----------------------+
|    ,---.        ,---. |
|   (TN-1 )      ( TN-2)|
|    `-+-'  +---+  `-+-' |
|     +----|vR1|----+   |
|          ++--+       |
|          |       +-+----+
|          |       /Internet\ One path via
|        +-------+ Gateway  +---------------------+
|                \       /   Internet          \
|                 +-+----+                       \
+----------------------+                          \
                                                   \
+----------------------+            native traffic  \
|    ,---.        ,---. |            without encryption|
|   (TN-3 )      ( TN-4)|              |
|    `-+-'  +--+   `-+-' |                             |      +------
+
|      +----|vR|-----+   |                         +------+ CPE
|
|          ++-+       |                           |      +------
+
|          |      +-+----+                        |
|          |      / virtual\ One path via IPsec Tunnel |
|        +-------+ Gateway  +------------------------- +
|                \      /    Encrypted traffic over|
|                 +-+----+       public network    |
+----------------------+                           |
                                                   |
+----------------------+                           |
|    ,---.        ,---. |            Native traffic |
|   (TN-5 )      ( TN-6)|            without encryption |
|    `-+-'  +--+   `-+-' |            over secure network|
|      +----|vR|-----+   |                           |
|          ++-+       |                             |
|          |      +-+----+            +------+      |
|          |      /        \ Via Direct /customer\      |
|        +-------+ Gateway  +----------+ gateway  |-----+
|                \      /  Connect    \       /
|                 +-+----+            +------+
+----------------------+
```

Figure 2: Different Underlays to Reach Cloud DC

8. End-to-End Security Concerns for Data Flows

   When IPsec tunnels established from enterprise on-premises CPEs
   are terminated at the Cloud DC gateway where the workloads or
   applications are hosted, some enterprises have concerns regarding
   traffic to/from their workload being 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.

9. Requirements for Dynamic Cloud Data Center VPNs

   In order to address the aforementioned issues, any solution for
   enterprise VPNs that includes connectivity to dynamic workloads or
   applications in cloud data centers should satisfy a set of
   requirements:

   - The solution should allow enterprises to take advantage of the
      current state-of-the-art in VPN technology, in both traditional
      MPLS-based VPNs and IPsec-based VPNs (or any combination
      thereof) that run over the public Internet.
   - The solution should not require an enterprise to upgrade all
      their existing CPEs.
   - The solution should support scalable IPsec key management among
      all nodes involved in DC interconnect schemes.
   - The solution needs to support easy and fast, on-the-fly, VPN
      connections to dynamic workloads and applications in third
      party data centers, and easily allow these workloads to migrate
      both within a data center and between data centers.
   - Allow VPNs to provide bandwidth and other performance
      guarantees.
   - Be a cost-effective solution for enterprises to incorporate
      dynamic cloud-based applications and workloads into their
      existing VPN environment.

10. Security Considerations

The draft discusses security requirements as a part of the problem
space, particularly in sections 4, 5, and 8.

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.

IANA Considerations

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

11. References

11.1. Normative References

11.2. Informative References

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

[RFC8192] S. Hares, et al "Interface to Network Security Functions
          (I2NSF) Problem Statement and Use Cases", July 2017

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

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

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

   [RFC4664] L. Andersson and E. Rosen, "Framework for Layer 2 Virtual
             Private Networks (L2VPNs)", Sept 2006.

   [BGP-SDWAN] L. Dunbar, et al. "BGP Extension for SDWAN Overlay
             Networks", draft-dunbar-idr-bgp-sdwan-overlay-ext-03,
             work-in-progress, Nov 2018.

12. Acknowledgments

Authors' Addresses

    Linda Dunbar
    Huawei
    Email: Linda.Dunbar@huawei.com

    Andrew G. Malis
    Huawei
    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