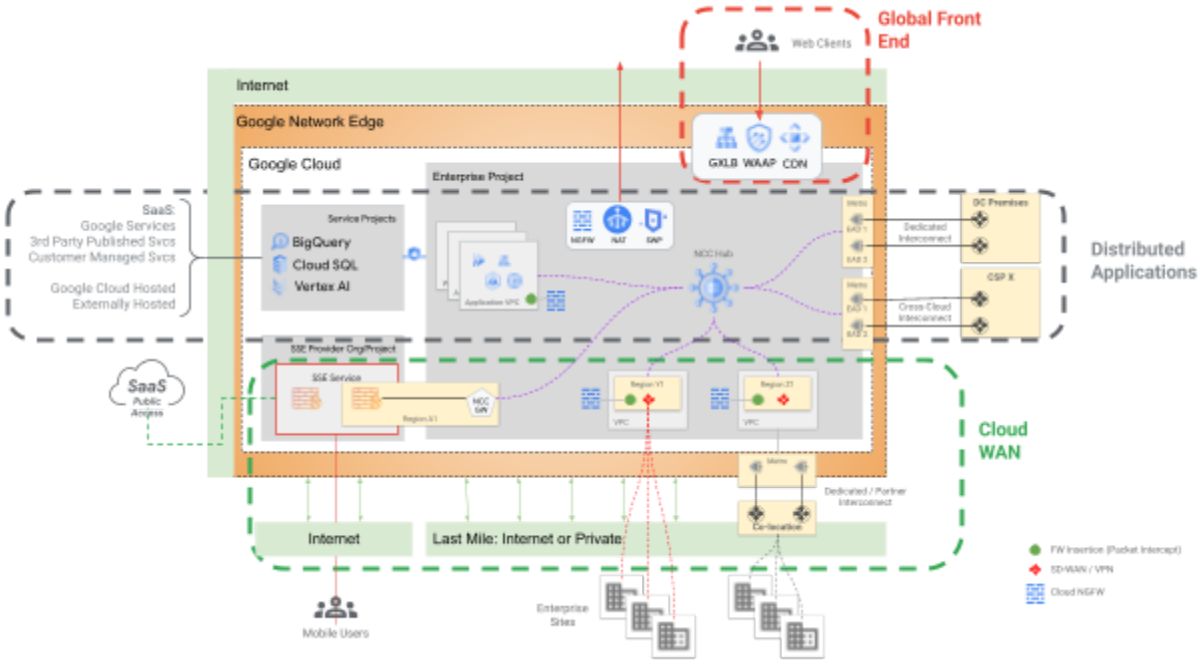
Google Cloud

Cloud WAN

Solution Deep Dive

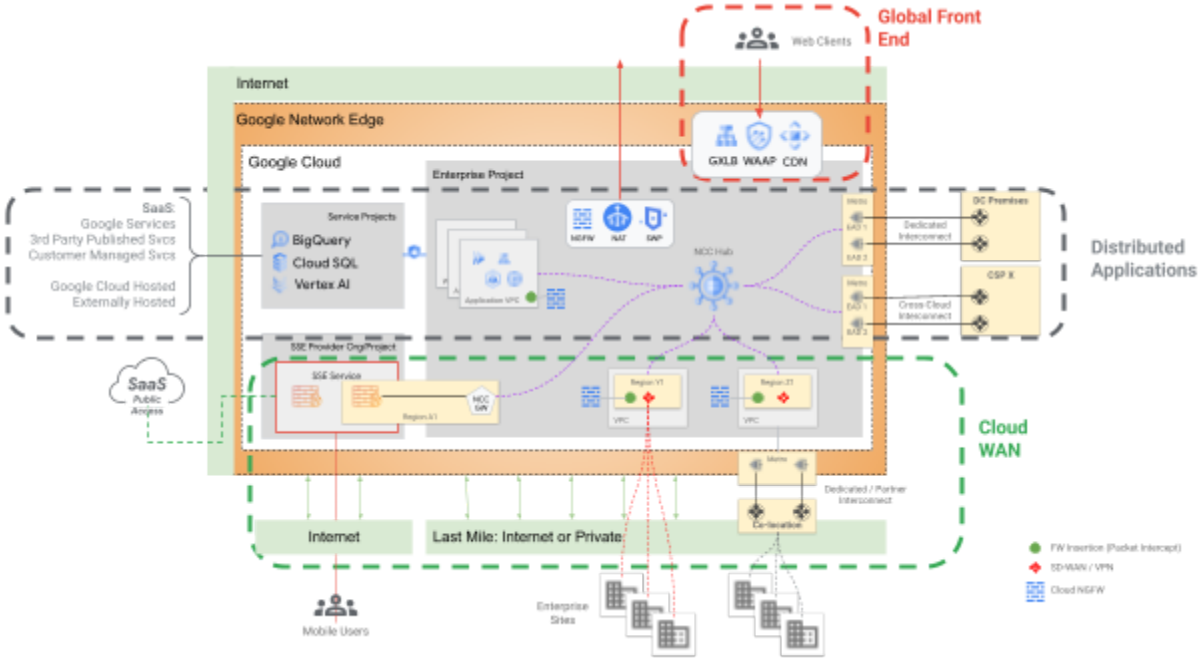


Global Connectivity, Simplified, and Secured

Get Google-level performance and security for your enterprise, with easy scaling and a cloud-native approach

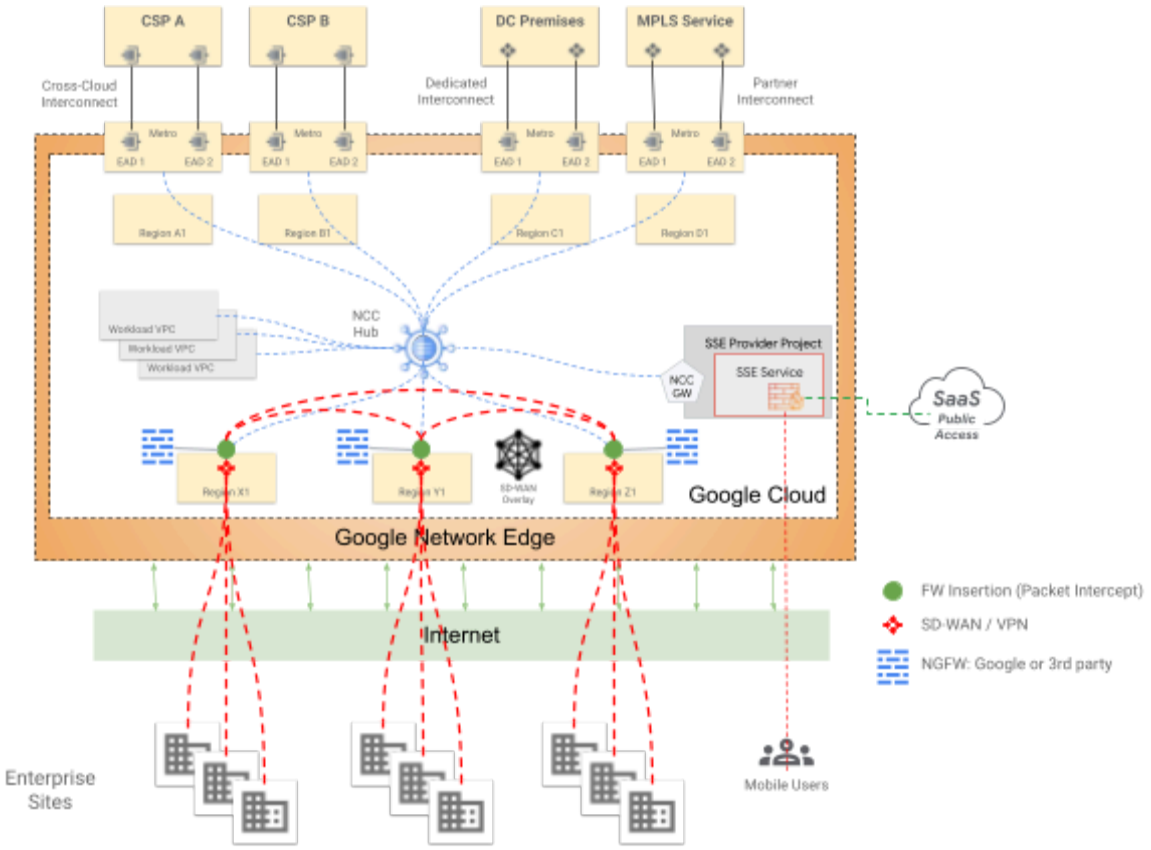
Cross-Cloud Network is a global cloud networking platform that offers any-to-any connectivity, enhanced application experience and ML-powered security for users and workloads wherever they may be. Cross-Cloud Network provides the infrastructure to interconnect the different networks at play in a global multicloud environment and optimally deploy security services, it is therefore the hub for connectivity, security and application delivery services in a multicloud environment. By centralizing these services through Cross-Cloud Network, enterprises can simplify the connectivity and security challenges of supporting multicloud applications, while optimizing the connectivity paths and performance that the applications require. Cross-Cloud Network enables the deployment of distributed applications across multiple clouds and private on-premises data centers, the private access of these application services for the hybrid workforce over Cloud WAN, and a global platform for web access to these applications via a Global Front End.

Cloud WAN is a cornerstone component of Cross-Cloud Network that enables the hybrid workforce to privately access the hybrid cloud data center and securely access public SaaS via cloud-hosted security service edge (SSE) services. The following diagram shows Cloud WAN in the context of Cross-Cloud Network.



This solution allows enterprises to use Google's global network as their private enterprise wide area network (WAN) backbone. With over 202 points of presence (PoPs) around the globe, customers can quickly and efficiently on-ramp traffic onto Google's global network with minimal latency in virtually any location in the world. Enterprises benefit from a global network built for mission-critical applications that offers high throughput, low latency, unprecedented path diversity, and high reliability through Google Cloud's premium Network Service Tier. Enterprises can consume this network as a service, and avoid most of the operational burden of running a wide area network backbone. A variety of connectivity options are available to connect to the Google network: Dedicated Interconnect and Partner Interconnect provide high-capacity private connectivity from customer premises with reliability SLA guarantees; Google's extensive internet peering infrastructure allows reliable connectivity via Verified Peering Partners for optimal SD-WAN connections; and native integrations of security service edge (SSE) services enable efficient on-ramp and off-ramp of mobile user traffic.

Along with enterprise class connectivity, Cloud WAN offers an ideal surface for the deployment of cloud-native security enforcement points. Traffic for mobile and on-prem users connecting to public and private applications and the internet, is secured with your choice of SSE service, Cloud Next Generation Firewall (NGFW), and/or 3rd party firewalls. The insertion of these security enforcement points is seamlessly enabled by packet interception using firewall policies without the need for complex routing exceptions. The result is an optimal combination of security enforcement in the cloud network data path and flexibility for users to choose their preferred security stack.



As depicted in the figure above, the solution provides a series of options for private IP connectivity from the customer premises to Google Cloud regions that are in close proximity to the customer premises. The VPC network across the different Google Cloud regions uses the Google Cloud backbone to deliver a global private IP network for the enterprise. This high performance end-to-end IP

connectivity is leveraged to provide connectivity between the different external networks in different regions, the google backbone is also leveraged as an underlay to deploy an SD-WAN overlay that enables granular routing and policies at scale for users connecting from company sites. The SD-WAN overlay head-ends form an overlay to the sites, but also with each other, to effectively form an overlay access network that is peered at every headend region with the VPC network in Google Cloud. All user traffic to and from private or public applications is steered through the VPC network where firewalls (Cloud NGFW or 3rd party) and/or SSE services can be inserted seamlessly in the data path based on policy.

Why use Google's Backbone?

Connect anywhere, anytime using Google Cloud's premium tier network

Google Cloud operates data centers in 42 strategic regions to bring cloud services within close proximity to industries in any geographic region around the world. This global footprint of data centers is interconnected by Google Cloud's planet-scale network which is accessible at over 202 network edge locations. Network edge locations are facilities at which the Google network peers with the internet, or offers private connectivity to external networks. Internet and private connectivity are the two main mechanisms by which last-mile connectivity is realized for access to enterprise backbones.

In addition to having a rich set of network edge locations, the Google network, with the premium Network Service Tier, enables anycast reachability over the internet for public prefixes hosted in Google Cloud. Anycast reachability of prefixes via a highly distributed footprint of network edge locations effectively reduces the latency introduced by the last-mile connection and allows traffic to get on the Google network as quickly as possible[1]. Traffic can then travel over the premium tier Google network across regions to destinations hosted within Google Cloud or hosted in other cloud service providers (CSPs) or other customer premises. Traffic in the premium tier network enjoys a 99.99% reliability SLA, and will travel on the Google network as close to its destination as possible (cold potato routing). By bringing all traffic to Google's premium tier network, the enterprise can effectively use the Google network as their global backbone.

Procuring connectivity to the Google network can be done quickly and virtually anywhere. Lead times for provisioning are typically much lower than traditional DIY provisioning timeframes as enterprises will be connecting to a pre-provisioned infrastructure. Since the infrastructure is available virtually anywhere, enterprises have the ability to extend their geographical footprint easily, quickly and cost effectively.

Google peers with most ISPs and offers high reliability peerings through the Verified Peering Provider (VPP) program. Pervasive IP anycast advertisements ensure that internet connections to Google follow the shortest path. For customers, this means that they can count on a global presence of reliable internet peerings without any provisioning overhead.

Customers can connect privately to the Google network using Google Cloud Interconnect in over 159 locations. Customers can choose to use a Dedicated Interconnect or a Partner Interconnect. Dedicated Interconnects require provisioning by customers and Google Cloud, while Partner Interconnects are sub-rate services from pre-provisioned circuits that can be consumed on-demand from the partner service provider.

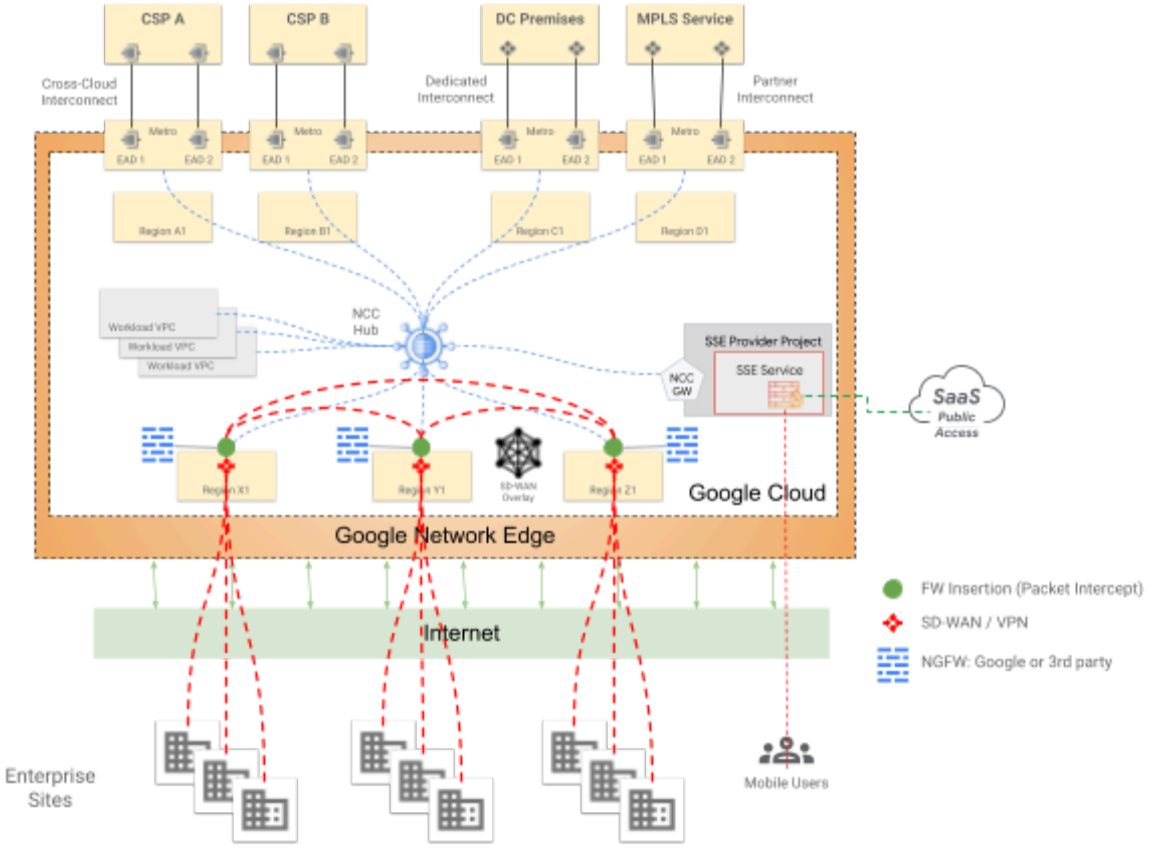
[1] There are exception cases in which the shortest last mile as determined per anycast routing doesn't necessarily provide the best latency. These are well known artifacts of the internet in very specific geographic regions.

The Google backbone runs on a very rich set of diverse paths. Google owns and manages a significant portion of the global subsea cable infrastructure, portions of which are dedicated to the Google network. This mix of shared and private subsea cables provides Google with a unique path diversity profile and makes the Google network uniquely resilient to outages that tend to have a global impact.

In addition to a rich set of cable paths, the Google network is implemented with router redundancy at each layer. Thus, the network has a very wide fan out of available paths that not only increase the total network capacity, but also allow the network to reconverge and route around failures more effectively.

The high path diversity in the Google network enables more advanced recovery mechanisms such as Protective ReRoute (PRR) to further increase the level of reliability by involving the hosts in making re-route decisions ahead of the network re-convergence process. With Protective ReRoute, the Google network is able to reconverge even in situations in which routing protocols and failure detection mechanisms break. The net effect is an increase in the measured reliability service level objective (SLO). Google has measured a reduction of up to 84% in the downtime in its production network as a consequence of the ubiquitous use of Protective ReRoute.

Reference Network Design



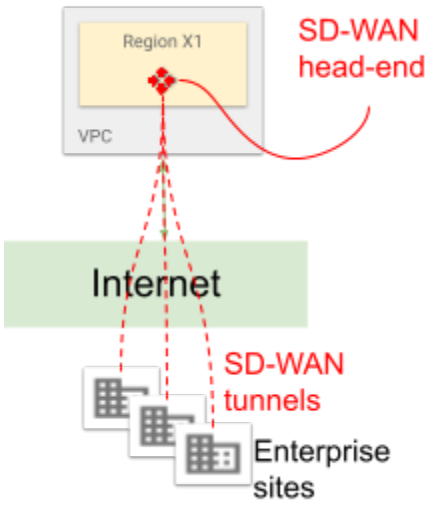
The reference network design for the use of the Google Cloud backbone provides high-capacity connectivity for users to connect to on-prem data centers and cloud data centers. Private connectivity for users at company locations is provided by integrating an SD-WAN overlay with Cloud WAN and

mobile users connect over a secure service access managed service that is natively integrated into Cloud WAN with fabric speed off-ramps. The reference network design comprises:

- A series of access network connectivity modules to connect the different types of external networks to the Google backbone in a multitude of regions.
- Wire-rate private connectivity between the different access connectivity modules.
- An SD-WAN overlay network deployed on top of the access modules and VPC topology to enable end-user private connectivity.
- Private access to SSE managed services for secure mobile user connectivity to applications and SaaS on the public internet

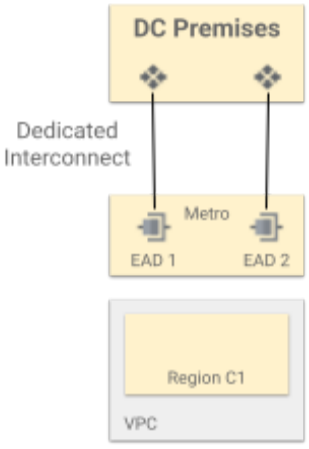
Access network connectivity-modules

There are different types of connectivity mechanisms with varying levels of performance and cost. Each of these connectivity mechanisms is deployed in a pattern. We refer to these patterns as connectivity modules. Several instances of these modules may be used in a backbone design to enable the appropriate connectivity for the different customer networks that will connect over the backbone. The different access modules include:

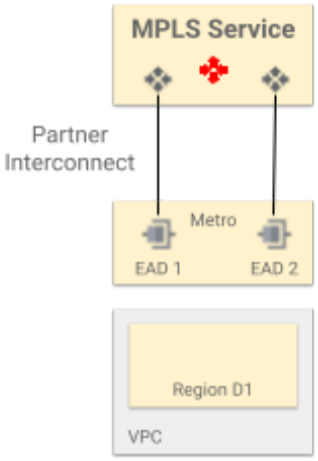


Site access aggregation: Sites will connect over the internet using an SD-WAN VPN to maintain privacy and enable the use of private address space. Google Cloud is ubiquitously connected to the internet in over 202 global peering points and offers highly reliable peering connections through its Validated Peering Partner program.

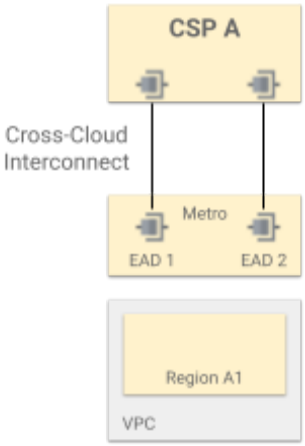
DC-premises connectivity: Private data centers connect using Dedicated Interconnect to provide a connection that can be MACsec encrypted, offers multiples of 10 and 100 Gbps connection speeds and is backed by availability SLAs of up to 99.99% uptime. This module allows on-prem data centers to connect to the VPC network.

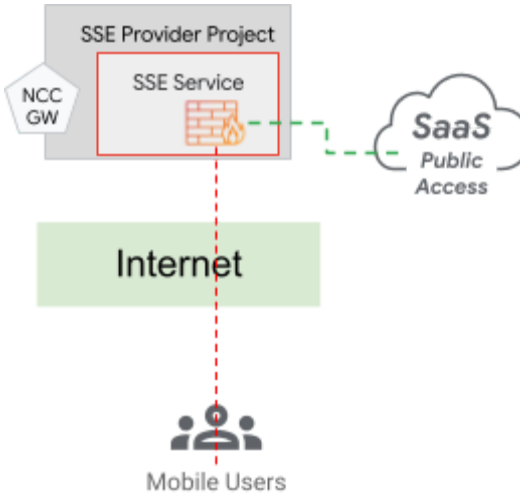


Existing enterprise private network connectivity: Enterprises may have an existing carrier enabled network, such as an multiprotocol label switching (MPLS)-based IP service. Interoperability must be established to this network to enable the phased migration to the new network design. Partner Interconnect is designed to privately connect the enterprise private network to Google Cloud. Partner Interconnect is a managed service offered by partner service providers to bring IP VPN services (such as MPLS) into Google Cloud. Full-rate or sub-rate 10 Gbps connections are available with Partner Interconnect with SLAs up to 99.99% availability. This module allows the MPLS network to connect to the VPC network; optionally, it may also enable reachability to a regional SD-WAN node.



Connectivity to other cloud providers: Google Cloud enables the direct connection to other cloud providers using Cross-Cloud Interconnect. The connections can be MACsec encrypted, offer multiples of 10 and 100 Gbps connection speeds and are backed by availability SLAs of up to 99.99% uptime. These connections are directly established by Google Cloud between itself and other CSPs so there are no assets to manage in co-locations, and the connections can be established across a wide range of location options based on the latency requirements between the different cloud providers. The peer CSP network is able to reach the Google Cloud VPC network via this module.





Mobile user connectivity and secure SaaS reachability: Mobile users connect to Cloud WAN via an SSE service. The enterprise has a choice of providers for this managed service through the Google Cloud Marketplace. On-ramp and off-ramp to and from the SSE service is provided natively in the VPC network using the NCC Gateway. The SSE integration also provides secure access to internet public SaaS for mobile users as well as users and resources connected through other modules in Cloud WAN.

Core Connectivity

Cross-Cloud Network (CCN) uses an any-to-any connectivity model that is orchestrated by Network Connectivity Center (NCC). Any resources that need to communicate with each other are registered as spokes of the NCC hub so that their connectivity can be orchestrated. NCC will then provision connectivity between the different spokes based on user intent and policy. The default connectivity model is a full mesh of connectivity, and custom topologies can be defined in the NCC policy environment.

External data centers connect to Cross-Cloud Network using the different patterns defined in the access modules. The patterns leverage high capacity interconnects such as Cloud Interconnect to connect on-prem data centers and Cross-Cloud Interconnect to connect to other cloud providers. Once connected to the Google network, these access network modules must be configured to communicate with each other. To achieve this connectivity, the hybrid connections in each access module are managed as hybrid spokes to a common NCC Hub. The NCC Hub will orchestrate a full mesh of connectivity between the different hybrid spokes. VPCs with workloads in Google Cloud can be added as VPC spokes to the same NCC Hub to enable connectivity between Google Cloud VPCs and external data centers. Cloud WAN adds the SD-WAN head-ends hosted in Google Cloud and NCC Gateways as spoke resources in the NCC hub to enable user connectivity to the Cross-Cloud Network.

User access connectivity regional considerations

The different site access aggregation modules will bring traffic to an SD-WAN node in a specific region. In order to interconnect these access modules (the access networks), the SD-WAN nodes are added as spokes of a common NCC hub. NCC will orchestrate the VPC networking to provide connectivity between its different spokes based on policy. The default policy allows all spokes added to the hub to communicate freely in a full mesh without restrictions. Connectivity can then be constrained to star topologies, if desired, and subnet and prefix distribution may be limited by policy, if required.

The collection of spokes managed by the NCC hub effectively forms a global software defined network with no forwarding choke points. All communications, whether intra-regional or cross-regional are distributed over a highly diverse multi-path network and will enjoy the high throughput, path diversity and resiliency that the Google backbone offers, without restriction.

To achieve the best availability and performance, it is recommended that multiple SD-WAN access modules be deployed in a diversity of regions. The selection of the regions in which to deploy the SD-WAN nodes and head-ends is based on the latency expected between a particular cloud data center region and the location of the enterprise premises to connect. Follow the best practices for region selection to arrive at the cloud regions with the best latency profile. For private data center premises and large campus locations, specific street addresses can be used to estimate latency to a particular cloud PoP. Consider evaluating low latency regions that minimize latency between the PoP and the region. The combined site-to-PoP and PoP-to-region latency can effectively be minimized to connect sites optimally.

The aggregation head-ends for sites connected over the internet will also have to be located in the best possible regions (usually those with the least latency to the sites). Google Cloud latency metrics for cloud regions to internet end points are available via the Network Intelligence Center's Performance Dashboard (View Google Cloud latency dashboard | Performance Dashboard). The dashboard presents median latency metrics aggregated across all of Google Cloud customer traffic as well as for each customer project traffic. This cloud-wide and per-customer-project metrics visibility enables customers to assess if any potential latency excursions are specific to their projects or across all Google Cloud

customer traffic, expediting troubleshooting. Similar considerations apply to the choice of region to deploy SSE nodes and the corresponding NCC Gateways that allow mobile users secure access to the data center assets in the Cross-Cloud Network.

In addition to the latency between Google Cloud regions and internet end points, Performance Dashboard also presents latency metrics for each possible zone pair in Google Cloud, at both Google Cloud-wide and per customer project scopes.

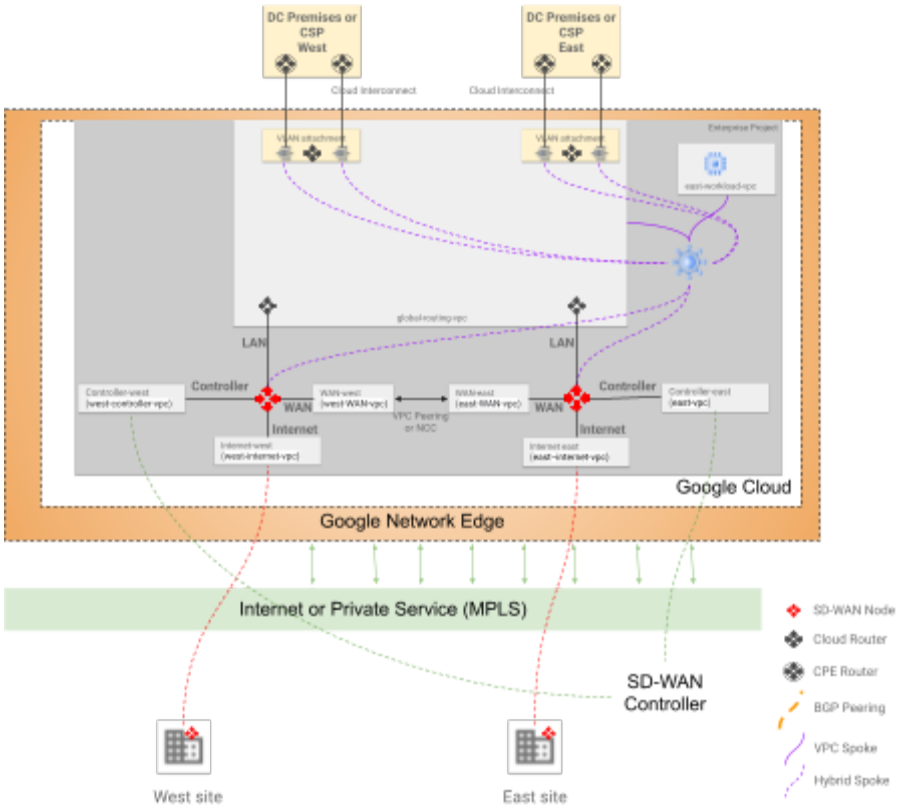
Once the regions are identified, an SD-WAN head-end can be created for each region.

SD-WAN router appliances in Google Cloud

The integration model for SD-WAN head-end appliances may vary slightly from one SD-WAN solution to another, but overall they are handled as multi-NIC router appliances and normally require the following interfaces:

- internet Interface: Terminate tunnels from branches/sites
- LAN Interface: Connects to the data center network (VPC network)
- WAN Interface: Used to tunnel traffic between head-end appliances
- Controller Interface: Connects to the SD-WAN controller

These interfaces are illustrated in the following diagram. Each interface must be in a different VPC. All LAN interfaces on different SD-WAN appliances must be in the same global VPC (with global routing mode enabled). The VLAN attachments for external data centers (other CSPs or on-prem) must also be in the same VPC. We will refer to this as the “routing VPC”, the use of this common VPC is necessary in order to support transitivity between external data centers and the SD-WAN network. Additionally, all hybrid resources (Cloud Interconnects and SD-WAN appliances) must be spokes of a common NCC management hub.



As depicted in the figure, the WAN interfaces may connect to different regional VPCs. In order to achieve the necessary underlay connectivity, these WAN VPCs must be peered to each other either with classic VPC peering or by NCC. Alternatively, a single global WAN VPC may be used.

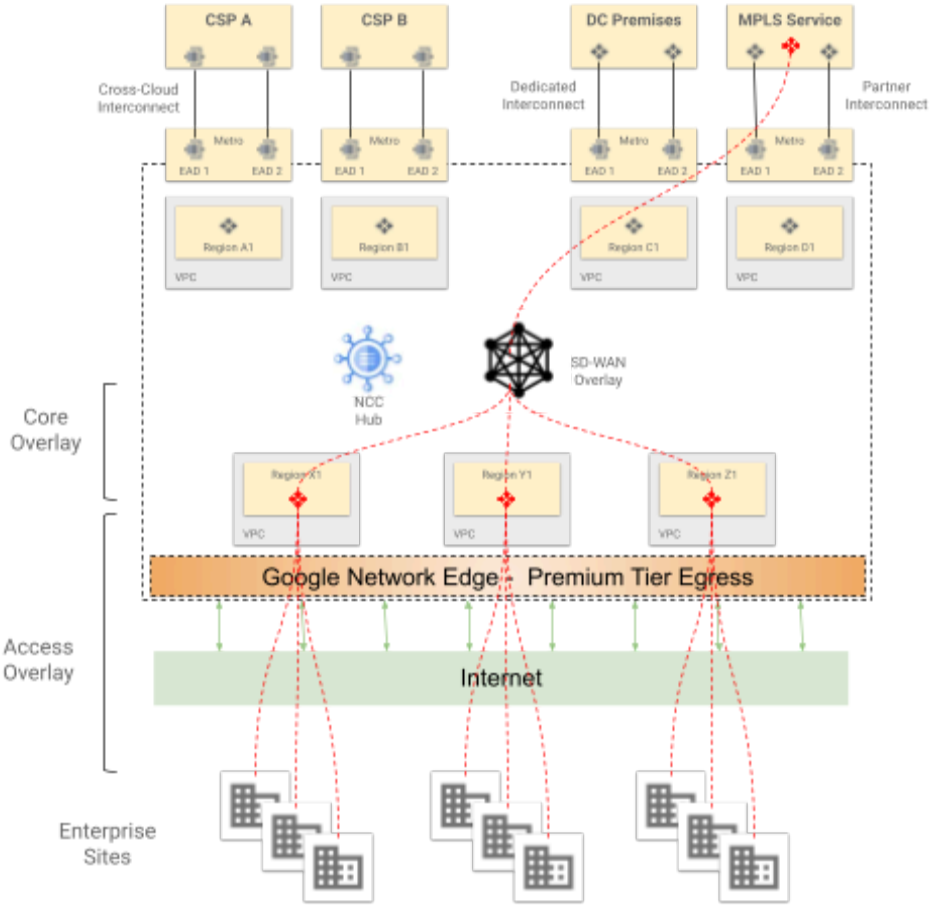
SD-WAN backbone overlay

Enterprises usually have a large number of non-aggregatable routes disseminated across their user sites. These sites can be connected to each other and to the VPC network using an SD-WAN solution. The SD-WAN would handle the large number of prefixes and announce summary routes to the VPC network at the different regions. To integrate an SD-WAN access network into Cross-Cloud Network, the SD-WAN head ends that will be deployed in Google Cloud will also need to be able to forward traffic to each other over tunnels. This is an SD-WAN tunnel overlay network running over the VPC network in Google Cloud. One way to look at this is to think of the VPC connections between SD-WAN nodes as the network underlay and the SD-WAN mesh as an overlay SDN network running on top of such an underlay.

Using an overlay network may enable an enterprise to circumvent regional restrictions on using the Google network natively for transport (this is limited by regulation in certain geographies). In addition, an enterprise may use the SD-WAN to selectively steer traffic over the internet or the Google network.

It is recommended that the overlay be structured with a core overlay that provides communication between the SD-WAN head-ends in the different regions and that an access overlay be deployed to aggregate traffic from the different user locations. The following diagram illustrates the structure of the overlay network. To minimize encapsulation overhead and maximize performance, the tunnels in the core overlay can be unencrypted IP tunnels (e.g. GRE) to save on encryption processing and overhead.

Since the tunnels travel over the Google backbone and within the enterprise's VPC network, encryption isn't necessary for the core overlay.



Managing large scale site routing

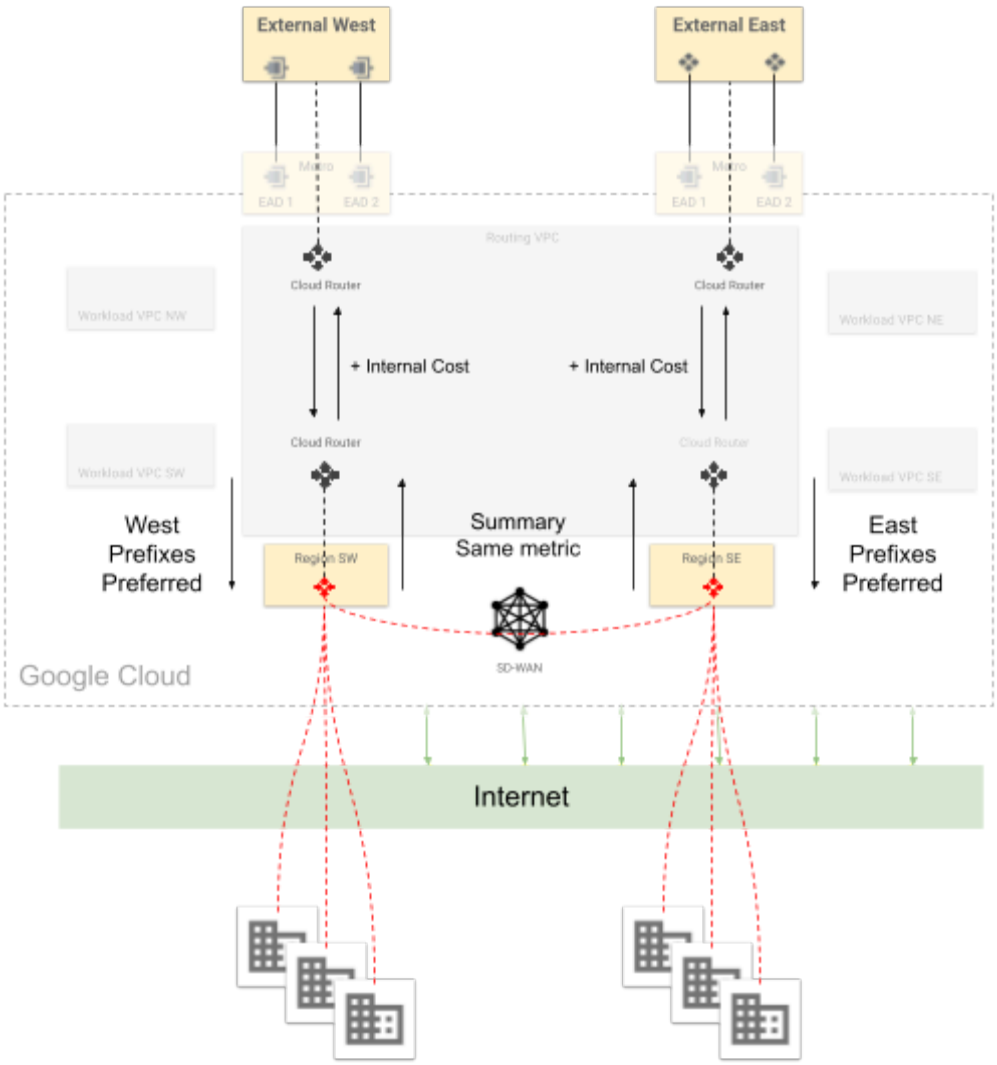
The number of specific routes dispersed across enterprise sites tends to be large and difficult (if not impossible) to aggregate. For enterprises looking to maintain a large number of non-aggregatable routes for site access, an overlay provides a scaling mechanism to handle these routes without advertising them into the Google Cloud network. The specific routes for enterprise private prefixes are handled in the SD-WAN overlay control plane. Route scale and policy capabilities are all realized in the control plane of the SD-WAN overlay network.

The following routing design enables a simple principle in which the SD-WAN network advertises a minimal set of covering prefixes into Google Cloud, the same set of prefixes is advertised from every SD-WAN head-end in every region. As these prefixes propagate across regions in the Google VPC network an internal cost is added to the routes. When the VPC network needs to forward traffic to an enterprise site, it will know which SD-WAN head end is the closest (by internal cost) and it will send the traffic to that closest SD-WAN head-end. Cross-Cloud Network is able to send the traffic to the closest SD-WAN head-end without the need to handle granular routing information. The SD-WAN will then handle the more granular routing. The actual destination site may or may not be directly connected to the selected SD-WAN head-end, but the SD-WAN control plane will have the more specific routing

information to get to the destination, and will tunnel the traffic to the authoritative SD-WAN head-end over the core overlay network leveraging the Google network as its underlay.

For the site-to-Cross-Cloud Network direction, the Cloud Routers in Google Cloud should advertise specific prefixes for the data center with metrics that reflect the proximity of those prefixes to the advertising cloud router. The VPC network adds these metrics automatically and the cloud router advertises the routes with the preference encoded in a border gateway control (BGP) multi-exit discriminator (MED) attribute. By having the SD-WAN network advertise the same covering prefix (with the same metric) everywhere and the Google Cloud network advertise specific preferences, the SD-WAN will act as a cold potato routing domain and traffic symmetry is achieved.

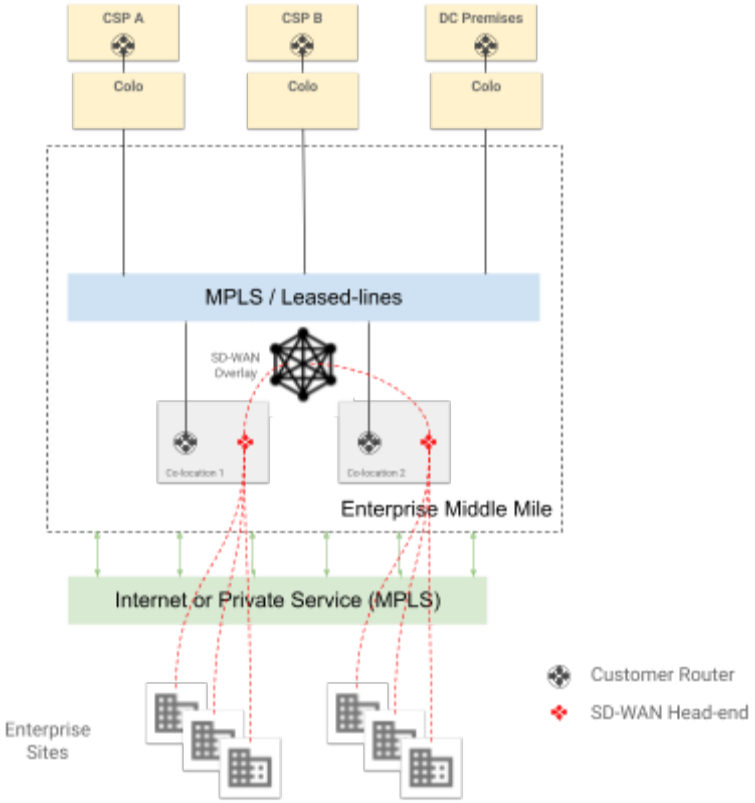
The advertisements and preferences are illustrated in the following figure.



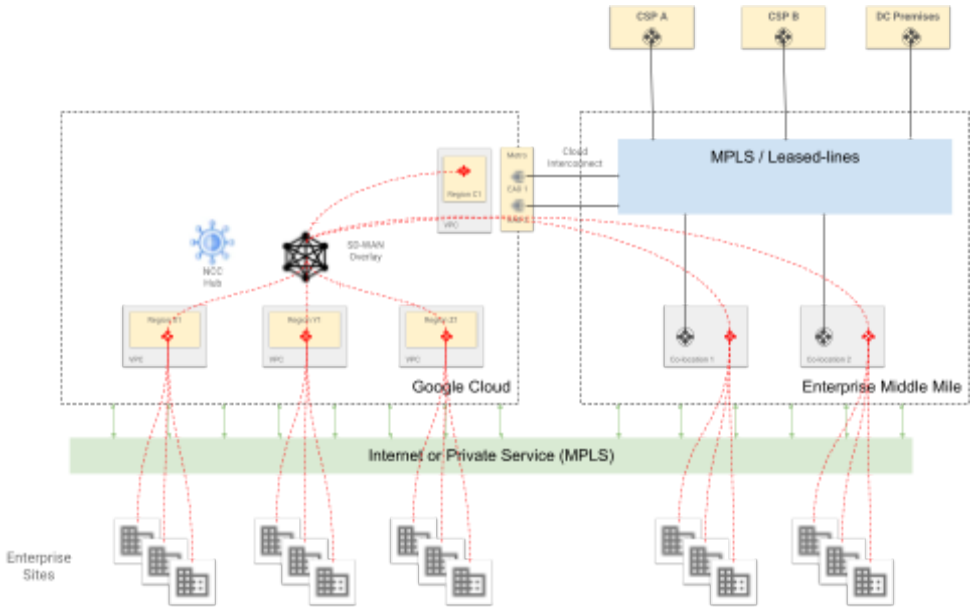
Interworking with existing enterprise WAN deployments

Many enterprises are actively deploying middle-mile networks, leveraging MPLS access services and deploying SD-WAN overlays. The typical deployment uses SD-WAN tunnels or an MPLS service to

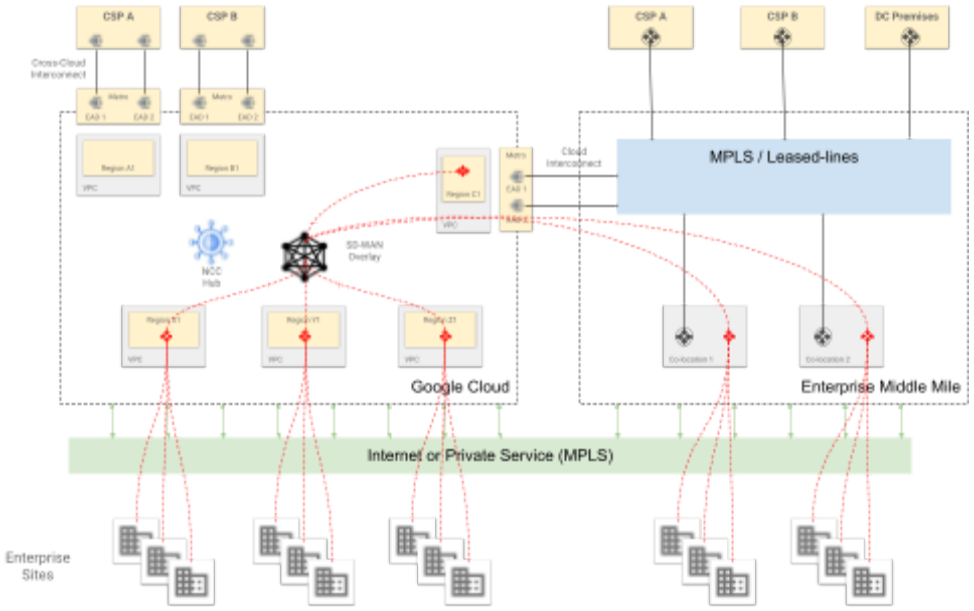
aggregate branch traffic at co-locations. The co-locations are interconnected to each other and to other data centers to form a middle-mile network as shown in the following diagram.



The existing middle-mile network can connect to the Google network using Cloud Interconnect (Dedicated Interconnect and/or Partner Interconnect) with up to 99.99% reliability SLAs. The SD-WAN overlay deployed over the middle-mile can be expanded to include head-ends hosted in cloud regions. Adding head-ends in Google Cloud, as shown in the following figure, allows the enterprise to easily expand their geographic footprint to aggregate user sites without operationalizing new co-locations.



Similarly, the enterprise may extend their network footprint to peer with other CSPs or on-premises data centers at locations that may offer better latency and performance. Since the SD-WAN network peers with the VPC network following the recommendations and routing outlined so far, the new connected networks do not need to connect directly to the SD-WAN overlay. This is shown in the diagram below.



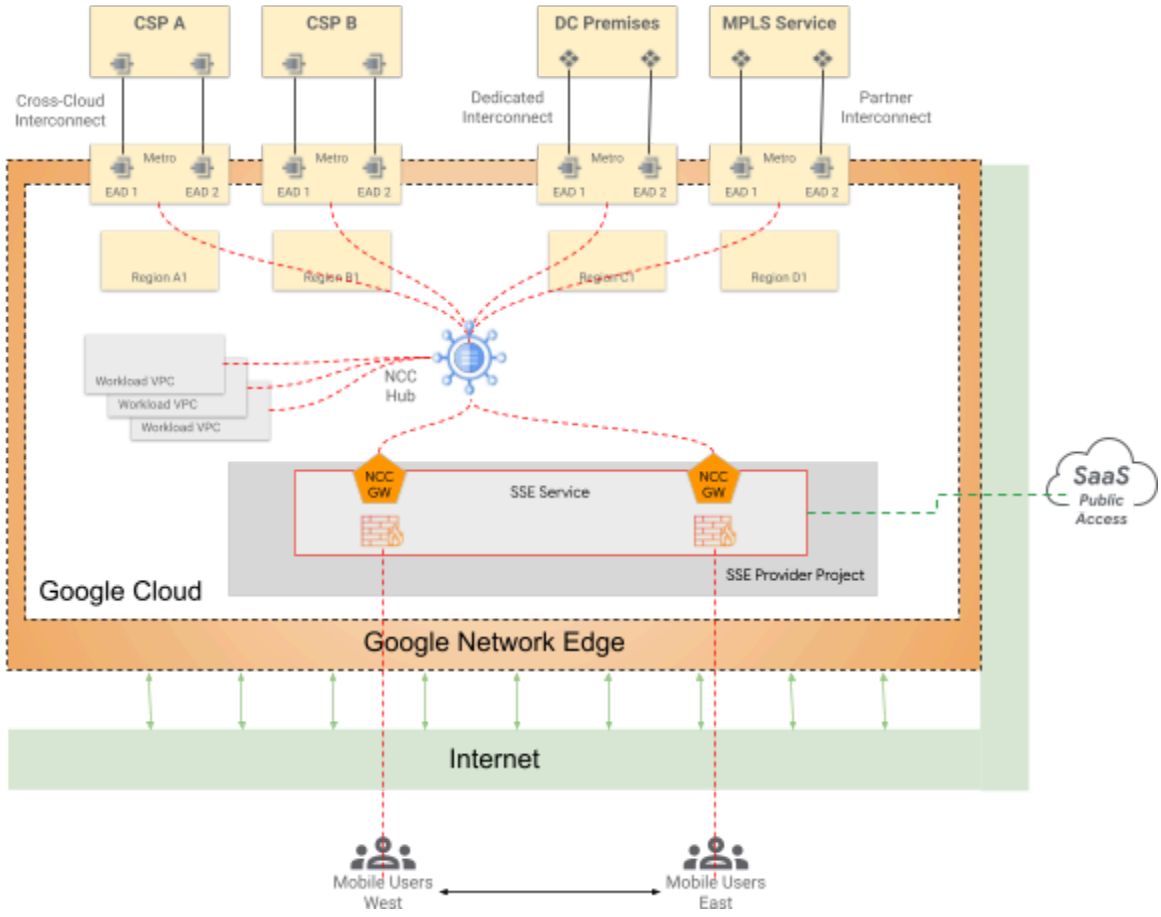
Secure User Connectivity

One of the main advantages of Cross-Cloud Network is the ability to reduce the number of security surfaces and consolidate the number of security stacks required by the enterprise to secure communications between application services, application users and application developers. Cloud WAN provides integrated security by consolidating security stacks into cloud-native form factors within Google Cloud. It offers managed security stacks as a service, handling scaling and lifecycle management. Cloud WAN secures traffic for both office-based and mobile users.

Secure connectivity for mobile users

Mobile users will connect over the internet to an SSE service using an encrypted connection from the mobile location to the closest SSE service node. Many of these SSE services are hosted in Google Cloud. Cloud WAN provides the necessary mechanisms to privately on-ramp and off-ramp traffic to and from Google Cloud hosted SSE services and the enterprise VPC network. These private connections are a significant improvement over previously necessary public connections that were subject to throughput limitations and added latency as they required traffic to be IPsec encrypted and sent over the internet. The NCC Gateway enables direct connectivity of the enterprise's SSE instance to provide better performance and a simplified deployment model. By using the NCC Gateway the SSE service can be handled as yet another spoke in the NCC scope of management and orchestration. Mobile user traffic connects to the VPC network after being processed by the SSE stack. Any other users (on-site) or workloads requiring secure internet access to SaaS can also be routed via the SSE stack by simple policy. The NCC Gateway spokes are depicted in the following diagram. As shown, NCC Gateway spokes may be deployed in different regions to enable traffic from the SSE service to be handed over to the VPC network at the preferred locations. Using this native integration of SSE services, mobile users can securely access workloads and private applications hosted in any data

center connected to the Cross-Cloud Network. Mobile users can also securely access public SaaS applications.



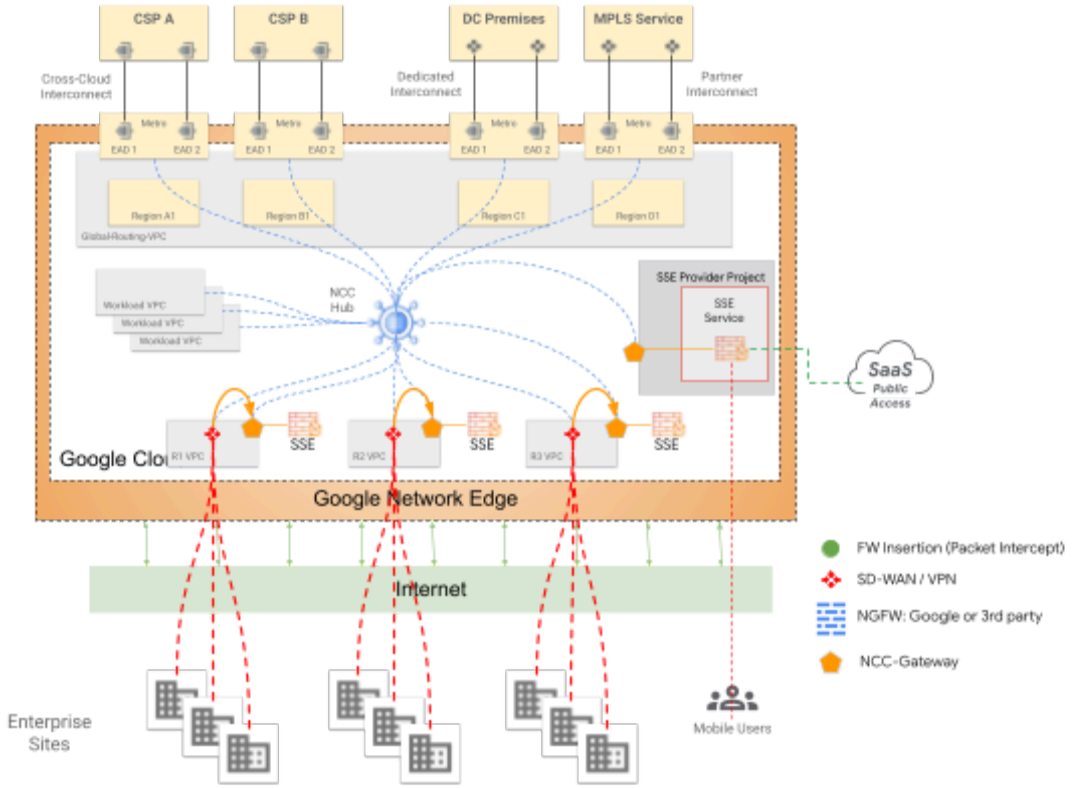
Secure connectivity for users connecting from a company site over SD-WAN

Users working from a company site connect over a Cloud Interconnect circuit or an SD-WAN tunnel to the Google network to reach application services and workloads hosted in Google Cloud, other CSPs or on-premises private data centers. The user traffic to and from a company site can be secured using one of the following security stacks:

- Security service edge (SSE)
- Next generation firewalls: Google provided or 3rd party

Secure users connecting from a company site with an SSE service for all destinations

User traffic to and from a company site connected using SD-WAN can be secured using an SSE service. This functionality is referred to as traffic on-ramp by some SSE providers. Traffic is steered to an SSE stack at each SD-WAN head-end by setting up the routing to use regional NCC Gateways as the next hop for forwarding, effectively inserting the SSE service in the path. All user traffic coming over the SD-WAN LAN interface is thus secured by the SSE stack. This allows users to securely access workloads and private applications hosted in Google Cloud, on-prem or other CSPs. It also enables secure connectivity to SaaS public applications reachable over the internet. The following figure illustrates this deployment pattern.



This pattern allows all site user flows to be secured via SSE. The following flows will be sent to the SSE service and then to their destination:

- Site users to workloads and private applications hosted in Google Cloud
- Site users to workloads and private applications hosted on-prem
- Site users to workloads and private applications hosted in other CSPs
- Site users egressing to the internet for access to public SaaS

Secure users connecting from a company site with NGFW for private applications and SSE to the internet

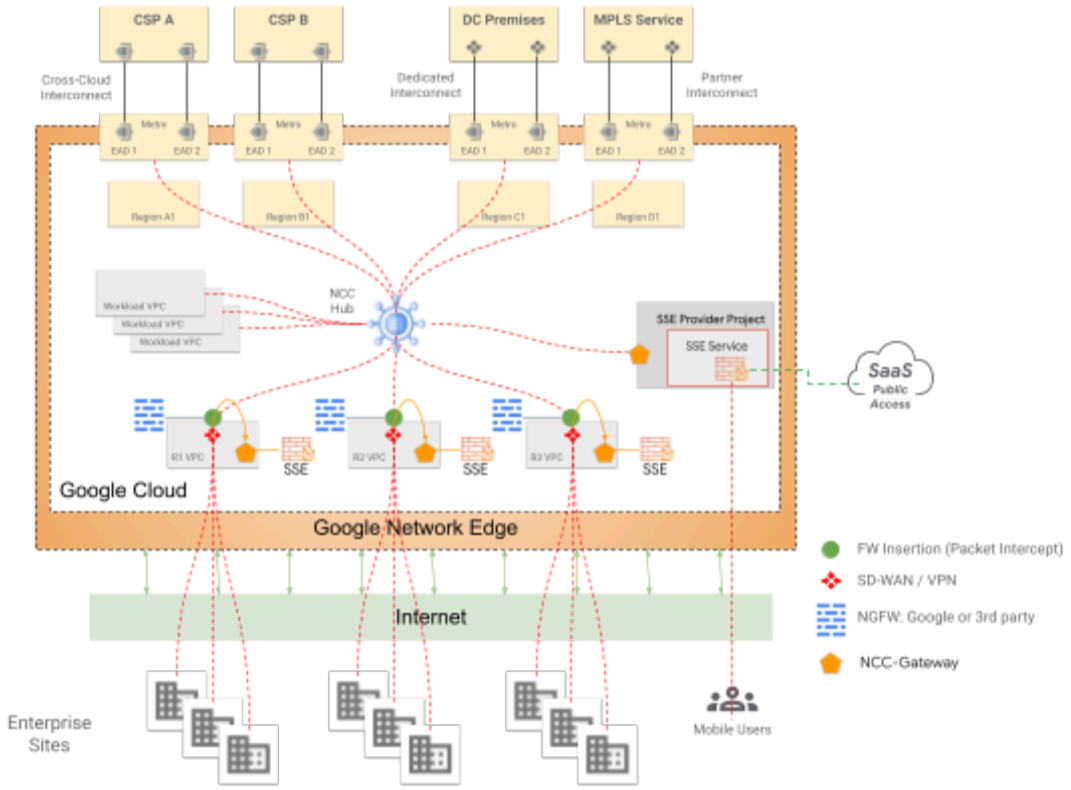
The user traffic to and from a company site connected using SD-WAN can be secured using the packet intercept technology that lets you place Cloud NGFW or third-party network appliances in the path of network traffic to or from workloads hosted in Google Cloud. In-band firewall options range from basic access controls to fully featured intrusion prevention systems (IPS) provided directly by Google Cloud or available from an ecosystem of partner security providers.

The following NGFW options are available:

- Cloud NGFW using firewall endpoints.
- 3rd party NGFWs using Google in-band Network Security Integration (NSI)

The interception policy can be enforced on the cloud-facing interfaces of the SD-WAN head-ends (LAN interfaces) and is effective for site user connections to any destination hosted in Google Cloud or in any of the data center networks connected to the Cross-Cloud Network. Using packet interception, traffic is subject to firewalling based on a traffic matching policy without the need for any routing

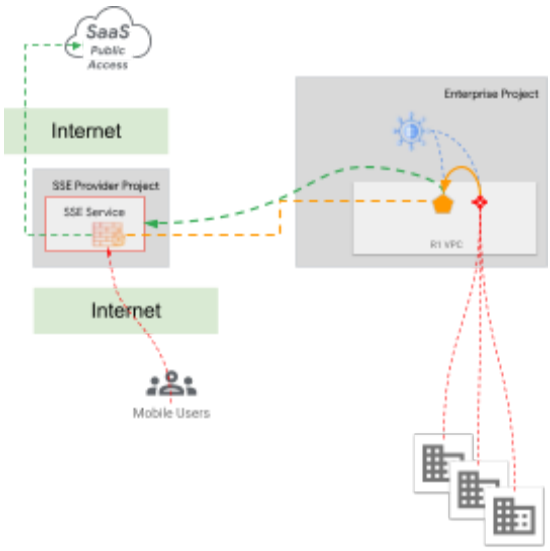
reconfiguration or traffic steering. The packet interception can be used to steer the traffic to Cloud NGFW firewall endpoints or 3rd party NGFWs using in-band Network Security Integration. The capacity and resiliency of the Cloud NGFW is fully managed by Google Cloud and offered as-a-service, while 3rd party NGFW appliances are managed by service producers (in-house or 3rd party) and offer the flexibility of using a 3rd party security implementation and tooling. The simplified insertion and cloud based lifecycle management make the insertion of security as a service in the Cross-Cloud Network efficient, simple to manage and cost effective. The following diagram illustrates the points of packet interception at which the firewall policies would be enforced for site traffic. This secures user access to private applications and workloads.



Internet bound site traffic aggregated using SD-WAN tunnels can be steered to an SSE service using private connectivity via the NCC Gateway. This is a high performance connectivity channel for traffic on-ramp to SSE services hosted in Google Cloud. The figure below depicts the design in which internet access to public SaaS is secured using an SSE stack. The following security surfaces are at play in this design:

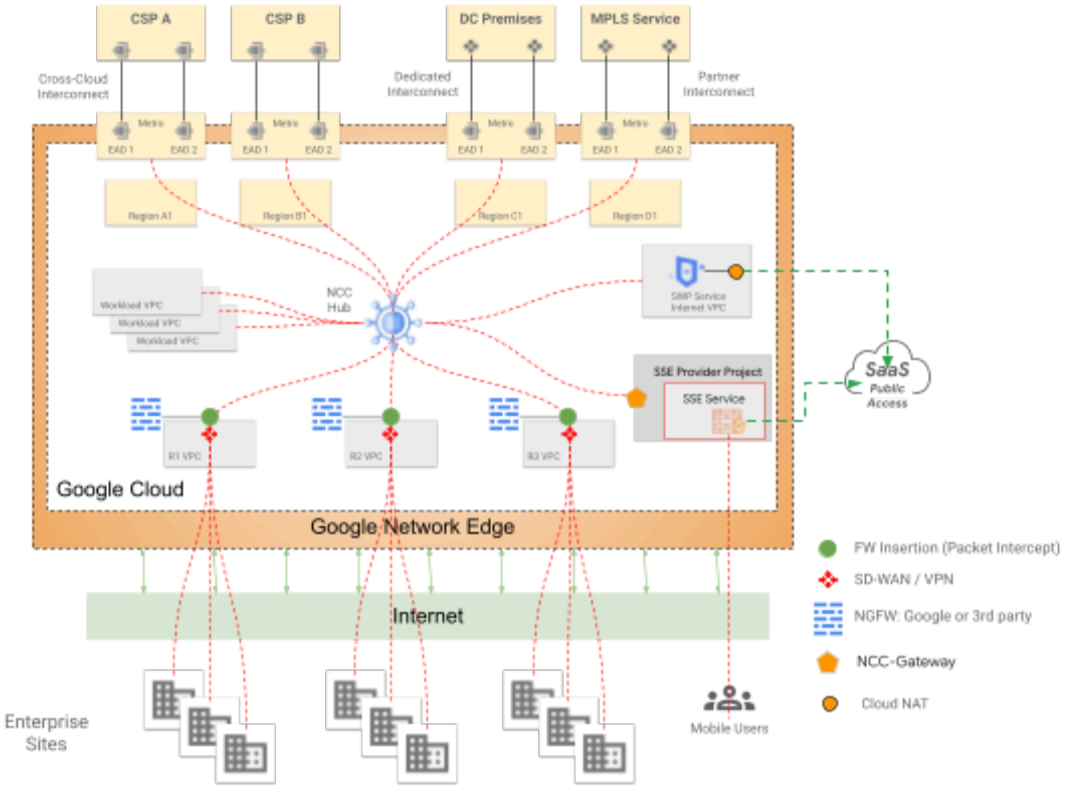
- All internet access is secured via the SSE stack
- Private app access from SD-WAN sites is secured using an NGFW (Google provided or 3rd party)

The internet access pattern via SSE is depicted in the following figure. An NCC Gateway in each region enables optimal distributed egress to the internet via SSE.



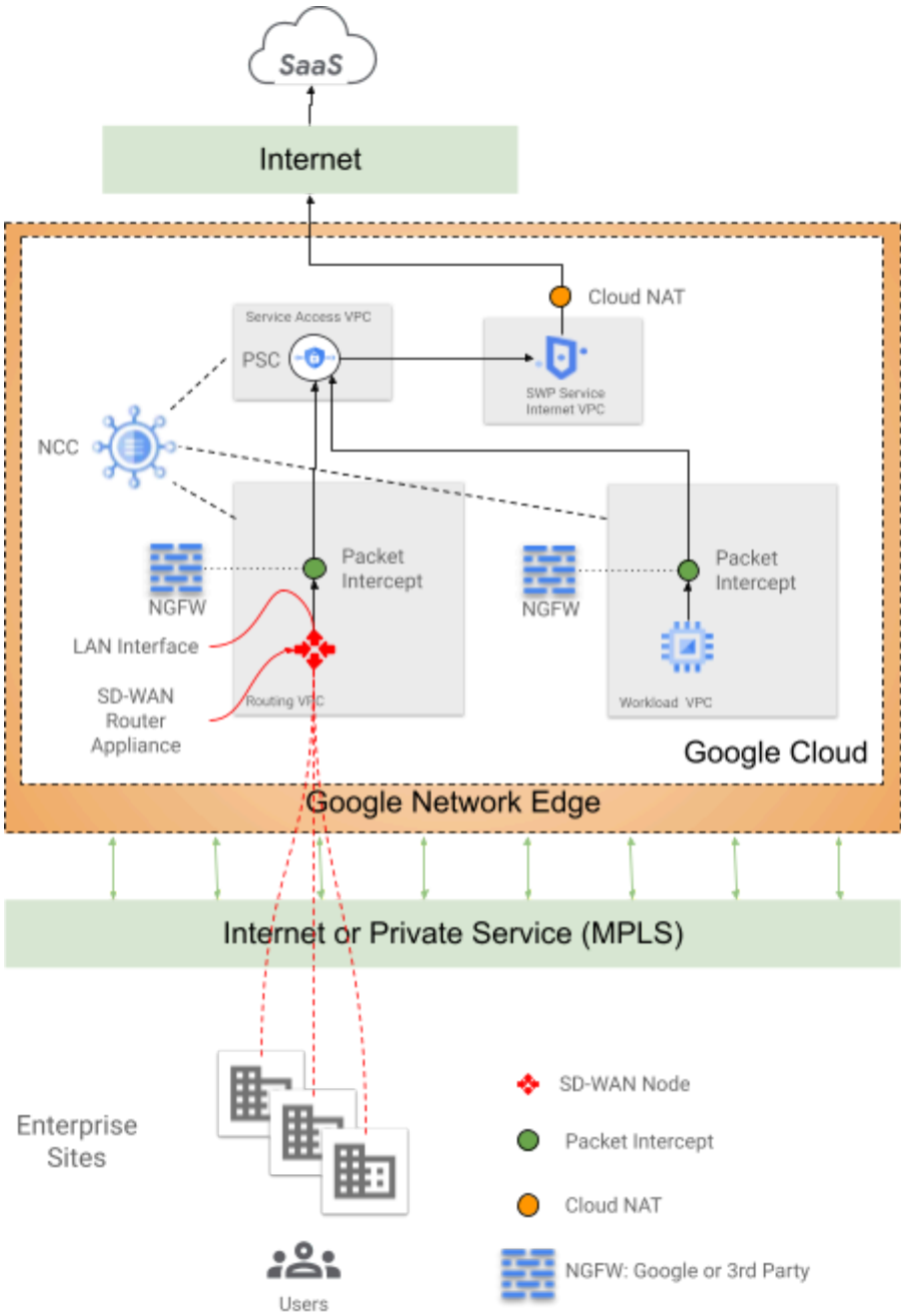
Secure users connecting from a company site with NGFW for all destinations (including the internet)

Another pattern customers could pursue is that in which all site user traffic is secured using NGFWs. This pattern leverages packet intercept to insert NGFWs in-band for access to private applications and workloads. The pattern also uses packet intercept in combination with Cloud NAT and Secure Web Proxy (SWP) to secure traffic destined to the internet.



Site traffic aggregated using VPNs such as SD-WAN tunnels can be secured using a combination of cloud-native network security mechanisms. Traffic aggregated by the SD-WAN headends can be

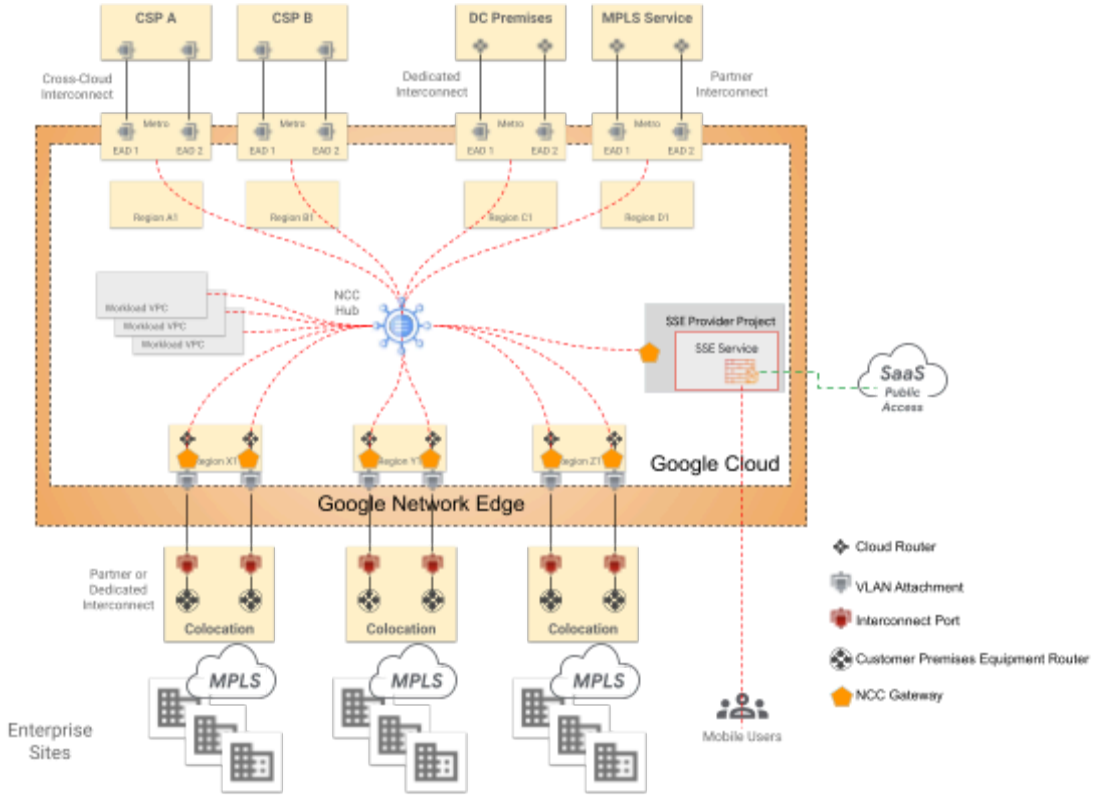
inspected and NATed at the VPC facing interface of the SD-WAN router appliances. The packet intercept technology in the Google Cloud network allows users to insert a Google provided NGFW or a 3rd party NGFW on all traffic leaving the SD-WAN appliance. The traffic can also be NATed at that point. Both firewalling and NATing are applied based on a matching policy for internet bound traffic. Routing for internet destinations may be set to go directly to the Google internet edge or it can be set to steer the traffic to regional instances of the SWP. For better control and URL filtering, it is recommended to steer the traffic via the SWP. The figure below depicts the design in which internet access to public SaaS is secured using a cloud-native security stack. As depicted, the same pattern can be used to secure internet egress connectivity for any workload hosted in Google Cloud.



Secure connectivity for users connecting from a company site over private last-mile networks without SD-WAN

Enterprises may have an established footprint of private last-mile connectivity, frequently implemented using an MPLS service. It may be desirable to leverage this existing access infrastructure as part of the Cloud WAN. These access networks can be peered directly with Cross-Cloud Network. There is a limit to the number of prefixes that can be exchanged with Cross-Cloud Network given by the capacity of the Cloud Router. To work within the capacity of Cloud Router, the MPLS CPEs may be able to summarize the prefix space in the access network. Exercise caution as the site address space may not be summarizable, in which case an SD-WAN solution may provide the necessary scale to integrate the access network into Cloud WAN. The different collocations with the MPLS site aggregation head-end routers will connect to the Cross-Cloud Network using either full-rate Dedicated Interconnects or fractional-rate Partner Interconnects. The MPLS CPEs will peer with Cloud Routers using standard BGP peering.

For these scenarios, it is recommended that all user site traffic be secured using an SSE service. NCC Gateways can be associated with each VLAN attachment over which the collocations are connected to Cross-Cloud Network. The NCC Gateways provide the insertion point to include the SSE stack in the flow of traffic. WIth this pattern, traffic destined to private applications/workloads as well as public SaaS applications on the internet can be protected with a consolidated stack. The pattern is illustrated in the following figure.



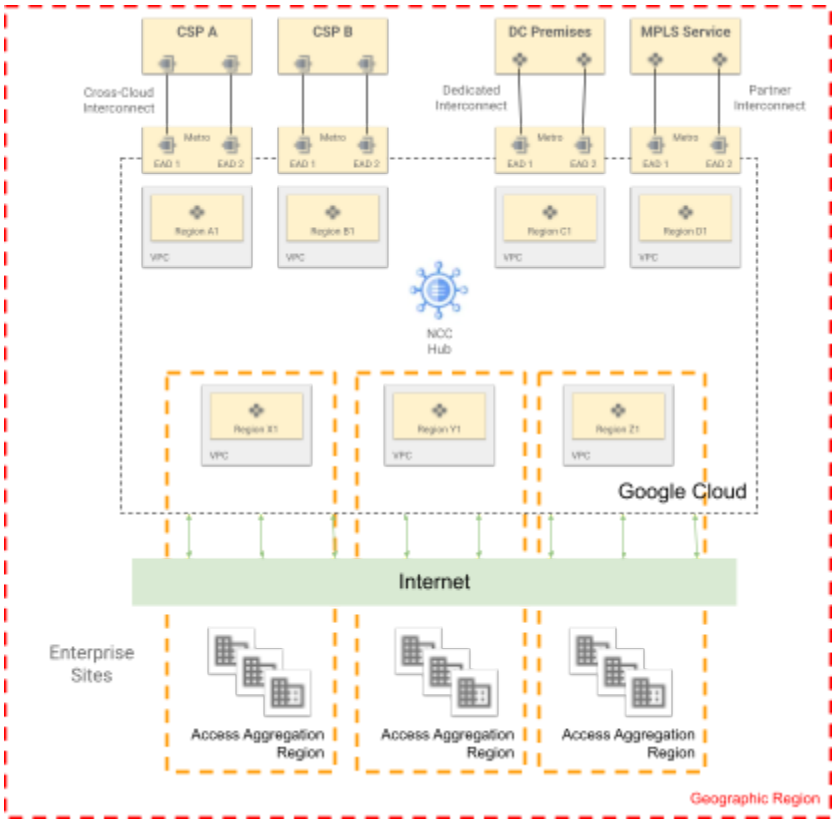
Regional Structure

The backbone is organized in geographic regions. A geographic region will include multiple Google Cloud regions within the same geographic area. For instance, a continent could define a geographic region.

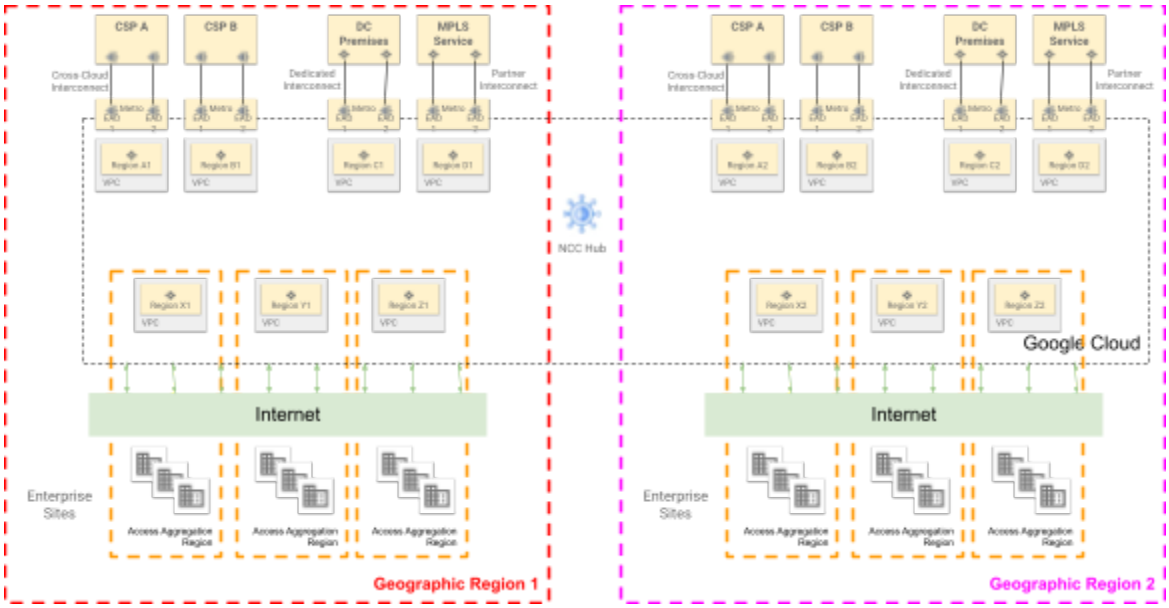
Each geographic region comprises multiple regions, each of which will host an access network aggregation point for a series of sites that are in close proximity. By deploying access network aggregation resources in different regions, the network aggregation layer can be scaled out horizontally and the latency between sites and the network aggregation nodes can be kept to a minimum.

In addition to the access aggregation, each geographic region may connect to the following external networks:

- One or more private data centers connected with Dedicated Interconnect
- Other Cloud Service Providers (CSPs) using Cross-Cloud Interconnect (CCI)
- The preexisting enterprise network backbone. These connections will use Partner Interconnect



A backbone may comprise multiple geographic regions. Connectivity between geographic regions is achieved over the Google network via NCC peering as all resources (VPCs, Cloud Interconnects, SD-WAN head-ends, and NCC Gateways) are spokes to the same NCC hub (even across geographic regions). Each geographic region is a fully-functional module that can be easily replicated.



Conclusion

The Google network offers ample benefits to enterprises that intend to use it as a transport to enable their global connectivity and SD-WAN overlays when these are utilized. Enterprises can enjoy network capacity, performance, and reliability in-line with what Google provides for its own planet-scale applications. A modular design streamlines the consumption of the Google network as a backbone and enables phased migration and interoperability with the existing enterprise network. Policy-driven security insertion allows the consolidation of security stacks in Cloud WAN with the SaaS benefits of lifecycle management, elastic scale and policy driven intent. Enterprises enjoy a menu of Cloud NGFW, 3rd party NGFW, and SSE services to protect their different security surfaces. Cloud WAN delivers performance, reliability, elasticity, security, and operational simplicity so that cloud practitioners can better support the infrastructure requirements from their development teams.