Cross-Cloud Network Solution Brief

Motivation

In the quest to leverage a curated mix of cloud application services, the diverse components of cloud applications are increasingly being distributed across a multitude of cloud and on-prem data centers. Companies seek to utilize the services they consider to be optimal for their applications and must therefore source these services from multiple cloud providers, for instance, an application may use AI services from Google Cloud, while maintaining business critical data on-premises and leveraging other application services from another cloud provider. This results in an application stack that is distributed across multiple data centers; communications between these components pose stringent privacy, latency, throughput and security requirements on an infrastructure that is inherently heterogeneous and disaggregated.

In recent years, the workforce has grown accustomed to a hybrid schedule in which work is done from a combination of private office locations and remote locations dispersed around the world. Office locations are connected over a network that offers privacy and a certain level of security assurances, as users move outside of the office, they connect over public networks which create a higher exposure to security threats. The business requires different levels of security for these different connection methods. Hence, not only has the workforce become more distributed, but the security stacks necessary to secure the workforce as they connect over public or private networks have become disaggregated.

The dispersion of applications and consumers across disparate and independent networks creates a significant challenge for businesses that find themselves having to engineer custom network backbones to interconnect the different data centers (on-prem and multi-cloud) and users; while steering different types of traffic through a variety of security stacks, each of which specializes in addressing the risk profile associated to a specific type of connection. As this trend evolves, a multitude of point solutions to specific issues emerge in the market and businesses find themselves tasked with integrating these diverse solutions in an attempt to keep up with the pace of the multi-cloud evolution. The result is a very complex and costly network infrastructure, which is inevitably sub-optimal and requires the business to coordinate a multitude of providers, technologies, business relationships and also train the operations personnel in the nuances of the different cloud environments. This situation has been somewhat tenable until now, however, the prevalence of AI powered applications has created an inflection point in which multi-cloud networking has become critical and can no longer be realized with the existing toolset and

models. A new multi-cloud networking paradigm would help achieve the required connectivity and security with the right agility, performance and cost to match the cloud standard.

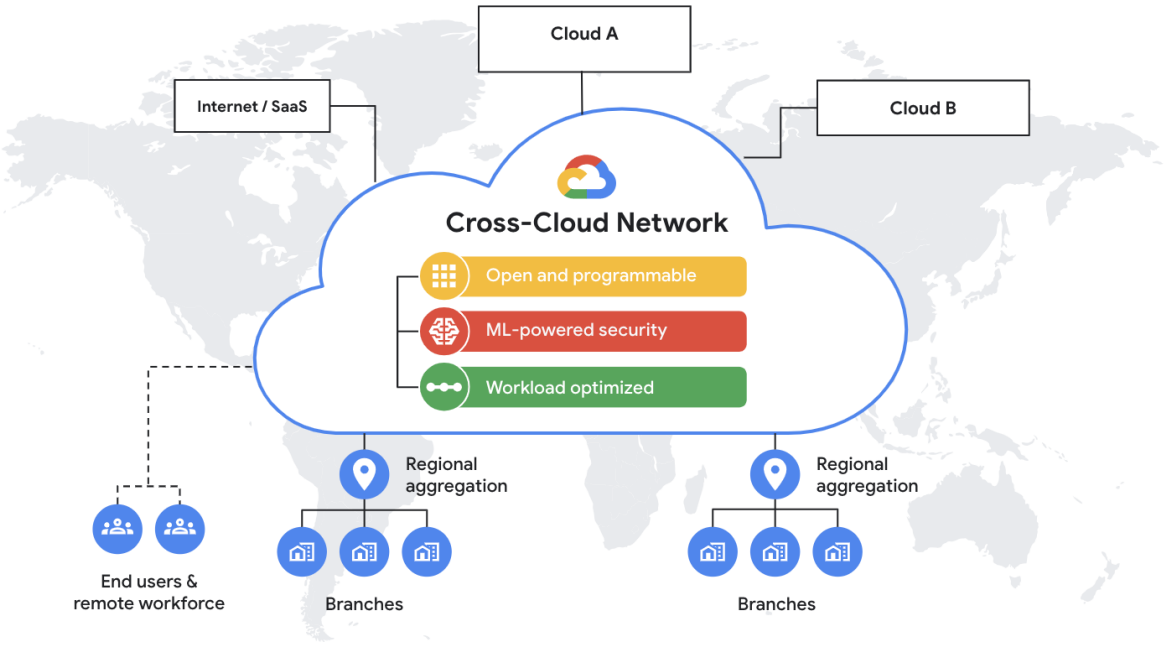
Solution

The Cross-Cloud Network is a global cloud networking platform with any-to-any connectivity, enhanced application experience and ML-powered security across all users and workloads wherever they may be.

The Cross-Cloud Network provides a fabric to interconnect the different networks at play in a global multi-cloud environment and optimally deploy security services, it is therefore the hub for connectivity, security and application delivery services in a multi-cloud environment. By centralizing these services through the Cross-Cloud Network, enterprises can simplify the connectivity and security challenges of supporting multi-cloud applications, while optimizing the connectivity paths and performance that the applications require.

Cross-Cloud Network addresses the multi-cloud challenges by delivering innovative services in three complementary areas:

- Planet scale any-to-any connectivity
- Enhanced application delivery
- Cloud native ML-powered security



Google Cloud

For more information visit **http://cloud.google.com/solutions/cross-cloud-network**

Planet scale any-to-any connectivity

Having a global presence can be challenging for enterprises as vendors, regulation, operations and pricing vary significantly across geographies, yet the distributed nature of business will often call for highly distributed connectivity in order to optimize the application experience and associated productivity. The Cross-Cloud Network enables the elastic use of the Google Cloud Network, to give the enterprise global presence, cloud grade performance and reliability.

With 187 Google Cloud Points of Presence (POPs) distributed across over 200 countries and territories worldwide, customers can use the Cross-Cloud Network to interconnect cloud networks and private locations elastically at any location.

Cross-Cloud Interconnect streamlines the connectivity between cloud providers by having Google Cloud handle the provisioning of multi-cloud high speed connections on behalf of the user. Cross-Cloud Interconnect provides a secure, high speed direct connection between Cloud Service Providers to guarantee privacy, throughput, availability and predictable latency. By using Cross-Cloud Interconnect, the customer does not need to maintain any infrastructure in co-locations or on-premises.

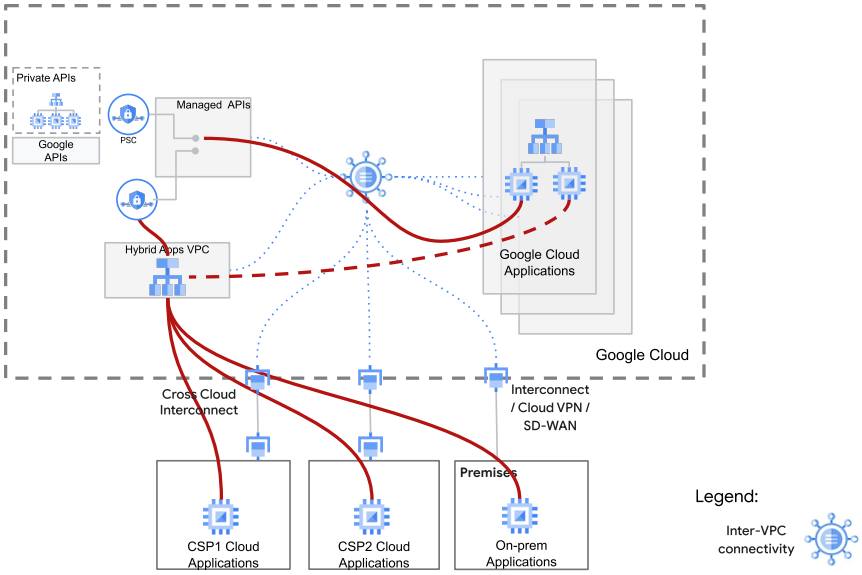
Private locations are easy to connect to the Cross-Cloud Network using any of the hybrid connectivity options available, which include: IPsec based HA-VPN tunnels, Partner Cloud Interconnect, Dedicated Cloud Interconnect or your choice of SD-WAN solution enabled by the Network Connectivity Center.

Enhanced application delivery

As applications modernize and migrate to the cloud, dependencies and preferences on specific cloud services develop. As a result different applications will be hosted in different cloud providers, furthermore some applications may be built with a combination of resources and services that are distributed across multiple clouds. This may be generally seen as an exercise in calling APIs across the different clouds, yet before these APIs can be called, proper connectivity channels must be established between the applications and their component resources and services so that every call doesn't travel out to the Internet and back into the cloud, which would be both insecure and inefficient. The Cross-Cloud Network enables multi-cloud communication between application components over optimal private connections by extending the capabilities used in Google Cloud to resources and services in other cloud providers.

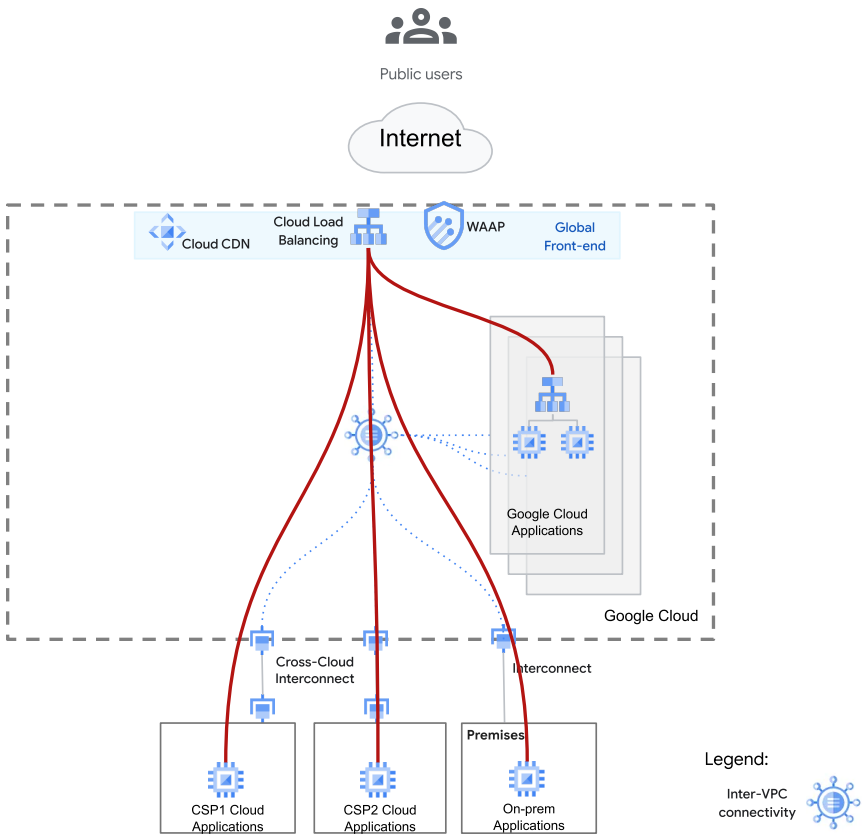
The Google Cloud Load Balancers provide mechanisms to group backend compute resources, balance the workload across the group members and achieve elastic capacity for different application components. The group of resources is reachable via the front-end IP of the Load Balancer, effectively encapsulating the group of resources as a single resource IP that can be used to assemble the application. Global Access to the load balancer front-ends allows the flexibility to include resource groups from any region in an application. Furthermore, with Global Backends, backend resources may be

distributed across regions, enabling multi-region resiliency and optimal load distribution across geographies. With the introduction of Hybrid Network End-point Groups, the backend resources can be reachable across a hybrid connection such as Cloud Interconnect or an HA-VPN. The functionality includes health checks that stretch into other networks over the hybrid connection. This effectively allows the user to model resources and services hosted in other clouds as IP end-points that are natively reachable in the Google Cloud. When the resource group needs to be consumed as an API, this abstraction can be refined further by using Private Service Connect to provide a private address for the resource group that can be accessed in the private network, across regions and/or over a hybrid connection. By having these abstractions, resources from disparate clouds and under disparate connectivity models can be brought together under a unified connectivity model that enables the effective assembly of an application across a multi-cloud environment.



Similarly, for public application access, the backend behind the Google Global Load Balancer (in the google global front end) can include a mix of resources hosted in GCP, on-premises or other cloud providers. This effectively allows customers to centralize their Global Front End for public facing applications with the Cross-Cloud Network, rather than maintaining different front ends with disparate models in different clouds. The Global Front End natively balances connections based on the geographic distribution of the demand to minimize latency and improve application experience. Rather than relying on elaborate DNS localization schemes, the Google Global Load Balancer offers a global anycast IP front-end coupled with connection localization. Traffic is sent to the global anycast IP from anywhere in the world and is promptly connected to the nearest instance of the global load balancer, which steers the connection to the closest relevant backend resources. Thus, connections are seamlessly optimized on a global basis for speed and latency. Also, applications using the Google Global Front End will have access

to Google's Cloud CDN to cache high demand static content and again improve application performance and user experience.



Cloud native ML-powered security services

As applications move to the cloud and the workforce turns to a hybrid model between the office and remote locations, the enterprise ends up managing a multitude of disjoint security stacks. Not only is this risky and complex, but it also requires serious compromises in the performance of the connectivity as traffic takes sub-optimal detours in order to be inspected by different security stacks.

With Google's Cross-Cloud Network, connectivity and the delivery model for the application are consolidated, paving the way for the consolidation of the security stack and the start of the journey to make these security services truly cloud native.

There are three coarse types of connection to be secured:

- Public access to public applications
- Employee access to applications (public and private)

Google Cloud

For more information visit **http://cloud.google.com/solutions/cross-cloud-network**

- Application to application communications

Public access to public applications

As previously mentioned, Google Cloud's Global Load Balancer provides the opportunity for users to unify the Global Cloud Front-end. Connections through this Global Front End must be secured. As part of the Global Cloud Front-end Google Cloud Armor provides a web application firewall (WAF) and distributed denial-of-service (DDoS) mitigation service. It helps protect websites and services hosted in any cloud from multiple types of threats, including:

- DDoS protection: Cloud Armor can mitigate both volumetric and targeted DDoS attacks.
- WAF protection: Cloud Armor can block a wide range of web application attacks, such as XSS, SQL injection, and denial of service attacks.
- Bot management: Cloud Armor can help you identify and block bots, like scrapers and spammers.

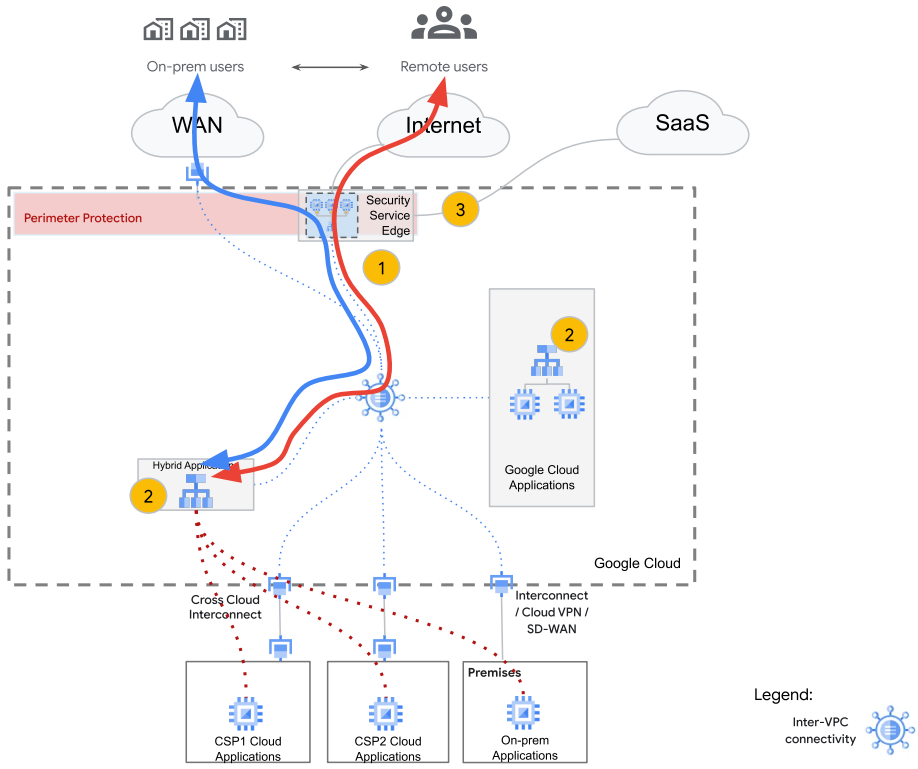
The Global Application Load Balancer offers an ecosystem of Service Extensions powered by an envoy proxy infrastructure. These Service Extensions are instrumental in delivering API security services and other protection critical to the Global Front-end.

Employee access to applications

Employees are a key vector for security attacks. As employees adopt a hybrid workstyle the risk is even greater. This risk has been managed by procuring a security stack that is application aware, assisted by artificial intelligence to recognize patterns and anomalies and rich in user and end-point authentication controls. Employees connecting over a public network to any resource must traverse such a security stack, the current industry term for such a stack is the Secure Services Edge (SSE). Beyond the threat and malware protections that a NGFW provides, SSE stacks include CASB, DLP and other identity verification services critical to securing connections over public networks like the Internet. There are many providers of SSE stacks and most offer the stack as a managed service. When offered as a managed service, the SSE stack is hosted at a series of locations around the world and reachable over the Internet, in order to maintain communications secure, traffic between the SSE stack and the applications must be encrypted and tunneled, reducing the effective throughput of the security stack significantly. Whether employees are connecting over the Internet or they are working from a company location, traffic must be steered through a security stack. Many of the SSE services pertinent to the public nature of the Internet may not be required when users connect over private links from company locations, this has resulted in enterprises managing separate stacks to secure employees working from the office vs. employees working outside the office. The challenge is therefore a combination of network complexity, performance and disparate security stacks.

The Cross-Cloud Network provides the necessary functionality to steer traffic for company users to a security stack that is hosted in Google Cloud. The security stack can be a fully managed SSE stack provided by Google's ecosystem partners, or it can be a more focused NGFW stack delivered as a 1P or a

3P cloud native service, or a combination of these options. Traffic can be steered according to policy to the appropriate security stack, making the routing to the SSE or NGFW seamless. Since the security stack is deployed natively in Google Cloud, there are no tunnels or encryption required, effectively bringing the performance of the stack to its full potential. The security stack also enjoys the benefits of elastic capacity as it is organized as a backend behind the cloud native load balancers. As traffic is centralized, the security stack can be consolidated. With Cross-Cloud Network employee security controls are more robust, simpler and don't impact application performance.



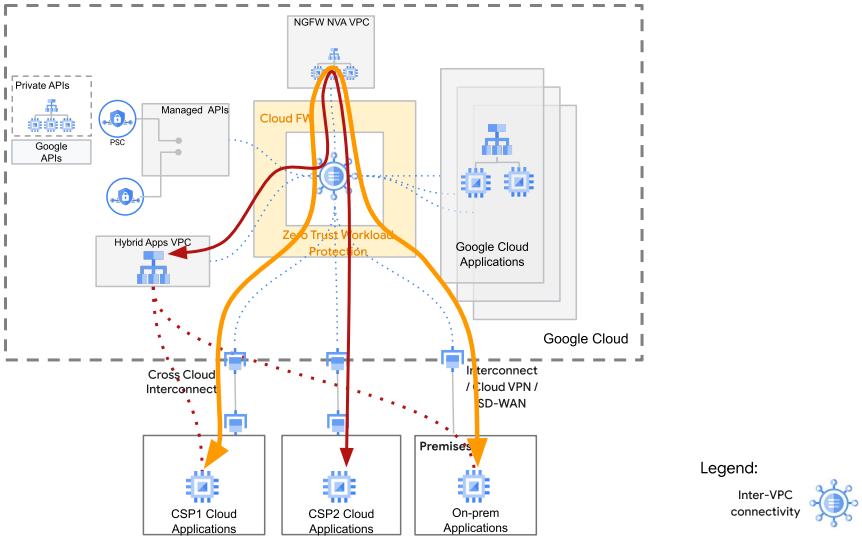
Application to application communications

As we assemble applications in a cloud native manner and normalize the model to that used in Google Cloud, applying security between the application components is now easier. The Cloud-native Next Generation Firewall (Cloud NGFW) can be used to apply Next Generation Firewall level inspection to the application flows that happen within Google Cloud and also with other clouds and on-premises data centers. The insertion of the Firewall controls is cloud native, the enforcement of the policies is done by firewall engines provided by our ecosystem partners. The NGFW service may be fully managed (Cloud NGFW), in which case the policies are authored as an extension to the Cloud FW policies in the VPCs; or it can be a self-managed NGFW, in which case the deployment is orchestrated by the Network Services Integration Manager and users can access the vendor specific policy console for authoring of the policy.

Google Cloud

For more information visit **http://cloud.google.com/solutions/cross-cloud-network**

For full coverage of all possible connectivity flows, a third party NGFW Network Virtual Appliance (NVA) can be inserted in the traffic path leveraging advanced routing functionality in Google Cloud. Policy Based Routing (PBR) enhances the functionality of the VPC routing stack to allow for policy based insertion of the security stack without the modification of the VPC connectivity design.



Google Cloud

For more information visit **http://cloud.google.com/solutions/cross-cloud-network**