Reblaze

Securely Using Google Cloud Platform

Reblaze Technologies LTD

Introduction

Cyberthreats are more numerous and varied today than ever before. Data breaches are common; Have I Been Pwned lists over 9.5 billion accounts that have been compromised. Companies that are breached suffer not only from loss of reputation and customer badwill, but also face potentially ruinous fines thanks to regulations such as GDPR.

DDoS attacks cause service disruptions and loss of revenue. Data theft and scraping can erode competitive advantages. Malicious bots wage credential attacks, deny inventory, and commit credit card fraud. The list goes on.

Cloud computing adds a new dimension to this situation. As companies move to cloud, they are using compute and storage resources which are designed to be accessible to the Internet. If best practices are not followed, security incidents can result.

This white paper will discuss security while using Google Cloud Platform (GCP). If your organization has not (yet) migrated to GCP and is considering doing so, you might be interested in this article: Securely Migrating to the Cloud.

Cloud security is of course a large subject. After reading this white paper, you will probably want a deeper dive into one or more specific topics.

Google has published a large number of resources about security and many other subjects; a good place to start is the GCP YouTube channel. Its videos vary in their length and depth. Some are introductions or tutorials: for example, Migrating to GCP? First Things First: VPCs (Networking End to End). Others are recordings of presentations given at the annual Google Cloud Next events: for example, A Security Practitioners Guide to Best Practice GCP Security (Cloud Next '18).

Shared Responsibility

Gone are the days when most businesses had to own their server infrastructure. With ownership comes responsibility, and with responsibility comes additional work and increased costs. The advantage of delegating work is that the responsibility becomes shared.

One of the significant advantages of using a cloud provider such as GCP is that you can delegate not only the management and operation of all infrastructure and hardware, but also the responsibility of securing it. In a nutshell, the cloud provider secures its cloud, and you—as a cloud user—secures whatever you use inside that cloud.

However, this does not absolve you from securing your cloud resources. In Google's overview of its Security Model, the company says:

> "Google is committed to doing its part in keeping your projects secure, but security is a shared responsibility. We've provided capabilities you can use to keep your project secure."

Here are some examples:

- GCP provides the ability to set user permissions for projects. It is your responsibility to set the least-privileged access for users.
- GCP provides the ability to control external traffic into a VM instance. It is your responsibility to correctly define the traffic that is allowed.
- GCP provides the ability to run Virtual Machines (VMs) in the cloud. It is your responsibility to maintain whatever OS and applications you run on the VMs, keeping them up-to-date with the latest security patches.

Compliance

Google Cloud is compliant with a wide variety of security and privacy standards. Among the more prominent are these:

In the Google Cloud Compliance Resource Center, there is a broad list of compliance offerings and certifications, such as PCI DSS, HIPAA, COPPA, GDPR, and many others. These also are a shared responsibility. All the compliances listed here mean that you can use Google Cloud in a compliant way. In other words, there is nothing on the cloud provider side that would prevent you from getting certified.

Still, it is ultimately your job as a cloud user to ensure that your products and services are compliant. Even though the cloud provider is compliant, you do not automatically inherit that compliance. It only means that the underlying infrastructure is compliant, and it provides a baseline for GCP users to build upon as they seek to achieve compliance for their own products and services.

Managed Services

When discussing cloud services, the term “managed service” is used frequently. However, the level of management can vary widely.

Traditional hosting providers typically offered physical or virtual machines, networking, and perhaps managed backups, but nothing more. These are known as Infrastructure as a Service (IaaS). Such offerings also exist in the cloud. In GCP, there is Compute Engine, which provides virtual machines and offers both ephemeral (local) and persistent (network) disk storage, in addition to GPUs. Google secures the host machines, but the security of the operating system and anything running atop it are not Google’s responsibility. The user still has the sole responsibility for many tasks such as patching, hardening, intrusion detection, and others.

While IaaS offerings may provide an easy migration path to the cloud, or for legacy workloads, more significant advantages begin with Platform as a Service (PaaS) products. Here are some examples:

- Cloud SQL is a fully managed database service for multiple database engines. With Cloud SQL, you no longer need to know how to securely operate a database or even have such knowledge in-house.
- Google Kubernetes Engine (GKE) is a better way to use containers, compared to managing a Kubernetes cluster.
- There are many serverless platforms available in GCP: Cloud Run for running stateless containers; Cloud Functions for event-driven serverless functions; or App Engine, the earliest serverless platform.

Platforms abstract away much of the management. This enables smaller teams to build highly scalable and performant web applications, while shifting the security burden to the application layer and above.

At the far end of the "managed" spectrum, there is Software as a Service (SaaS). These are end products for which users have little to no management responsibility. One example in Google Cloud is G Suite, which provides productivity tools for businesses.

From IaaS to PaaS to SaaS, the shared responsibility shifts away from the user to the provider. However, security should always be top-of-mind for cloud users; responsibility for security is never fully shifted to the provider. Even a SaaS product like G Suite, designed for end users and consumers, still requires following best practices for security (e.g., using a password manager and two-factor authentication).

The remainder of this white paper will focus on the lower end of this spectrum, and will discuss security issues that must be considered when running workloads on Google Cloud Platform.

GCP Security Products

Google Cloud Platform offers over a hundred products, dozens of which are related to security. Let's look at some of the most commonly used ones, including several with compelling use cases.

Cloud Armor

Google Cloud Armor protects web services against distributed denial of service (DDoS) attacks. It can also help mitigate and protect against some of the most common types of web attacks, such as cross-site scripting (XSS) and SQL injection.

By itself, Cloud Armor provides rudimentary protection. For robust security, it still requires an admin to monitor traffic and create rules in response to attacks. This can be very difficult, especially during the stress of an attack.

A better approach is to add Reblaze onto GCP, which creates a 'security engine' for Cloud Armor. Reblaze provides a next-gen WAF, DoS and DDoS protection, bot management, scraping prevention, load balancing, and more. Reblaze offers precise ACL (access control list) capabilities, UEBA behavioral analysis, and Machine Learning for accurate and adaptive threat detection. Additionally, Reblaze is a managed service that is continuously updated with the latest threat signatures.

In use, Reblaze detects threats, and automatically updates Cloud Armor, which blocks hostile traffic at the edges. Reblaze and Google Cloud Armor together provide full, automated protection.

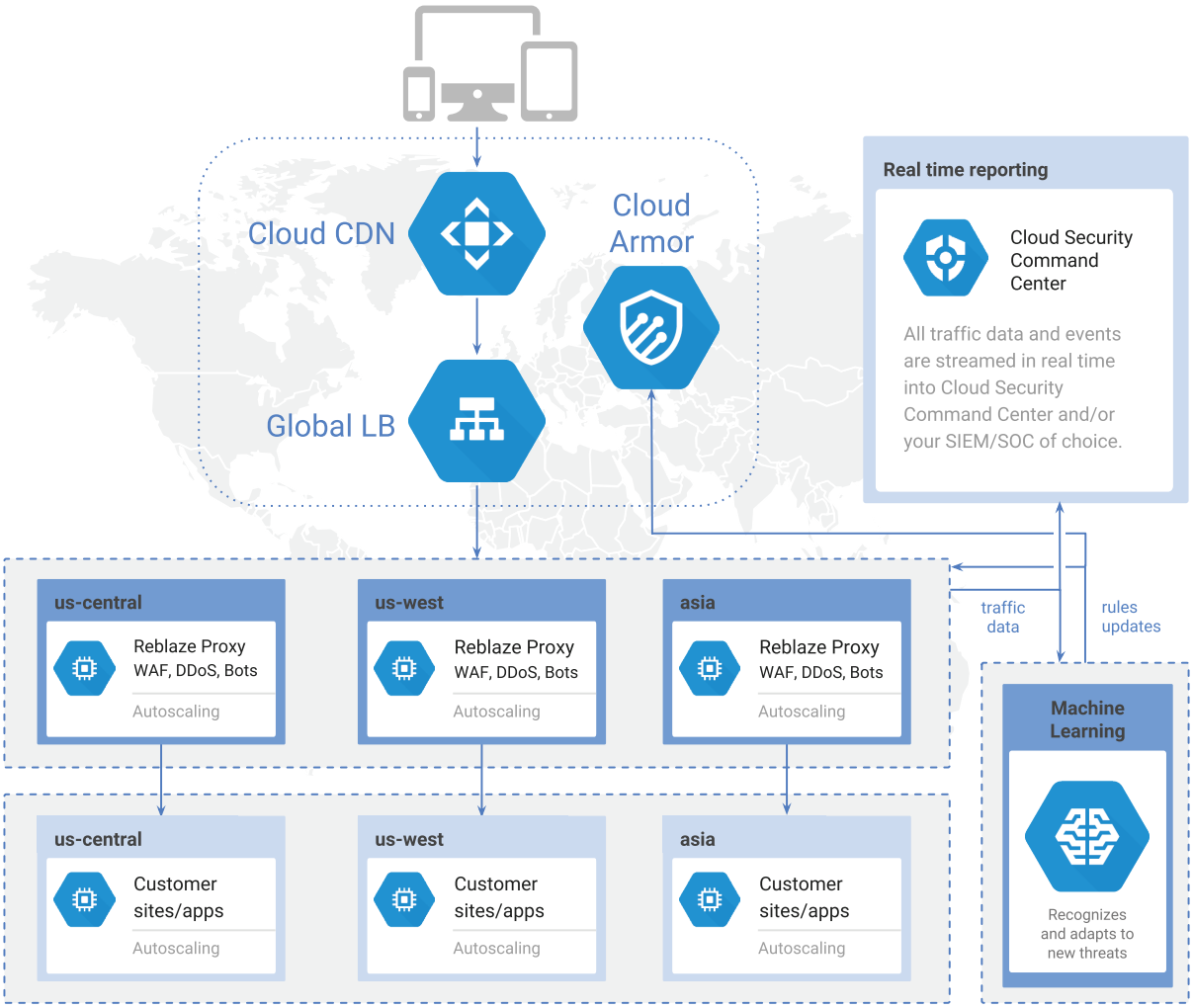
Cloud Load Balancing

Although this is not a security product per se, it is an important part of securing GCP workloads. Google Cloud HTTP(S) Load Balancing is implemented at the network edges in Google's points of presence around the world. Cloud Armor security policies are enforced at the load balancers, blocking hostile traffic before it can consume resources or access the upstream network.

Cloud Security Command Center

Cloud Security Command Center (CSCC) can work with many other GCP security products to provide visibility into GCP resources. For example, by integrating with Cloud Data Loss Prevention (DLP), CSCC can discover and alert you to sensitive or regulated data from a continuous scan of stored data. With the Event Threat Detection integration, it can also detect a plethora of threats from log analysis.

While this command center provides invaluable visibility, it does not automatically react to all threats. However, when using Reblaze, attacks are automatically blocked and reported directly to CSCC. You can read about this integration here: Cloud Security Command Center and Reblaze.



Reblaze running on Google Cloud Platform

Cloud Security Scanner

Cloud Security Scanner can automatically detect common web application vulnerabilities. The list includes XSS, mixed content, and outdated libraries or libraries with security vulnerabilities. This service is available for App Engine, Compute Engine, and GKE applications at no extra cost.

Cloud Audit Logs

Cloud Audit Logs provide highly secure, available, and immutable audit trails for user activity in Google Cloud. Today, most GCP services write audit logs; eventually, support is intended for all services. Cloud Audit Logs help provide a significant step towards compliance.

There are three different audit logs: Admin Activity, System Event, and Data Access. The first two are always-on, are free of charge, and provide a 400-day retention period. The Data Access logs are very high volume and require configuration before being enabled. Once enabled, Data Access logs are standard Stackdriver Logging logs, are billed as such, and they have the same 30-day retention period.

Honorable Mentions

A few other GCP products deserve a brief mention, as does the fact that encryption at rest is a default in Google Cloud. From keys that are fully managed by GCP, to customer-provided keys that Google Cloud uses in-memory and never stores, GCP is very flexible, depending on the security requirements.

The Cloud Identity-Aware Proxy (IAP) offers the ability to access network protected applications and VMs from untrusted networks without the use of a VPN. IAP works with both cloud and on-premise applications, and instances don't even require a public IP address. IAP is based on BeyondCorp, a zero-trust security model from Google, and is available for free.

Shielded VMs are hardened virtual machines that help protect sensitive workloads. Shielded VMs are automatically protected against boot-level and kernel-level malware and rootkits. They can also prevent data exfiltration by protecting secrets via a virtual trusted platform module (vTPM). The only requirement to enable these features is to choose a base VM image with Shielded VM support, which includes trusted UEFI firmware. There is no additional cost associated with Shielded VMs.

Resource Management

Proper management of GCP resources is essential. Just like web security, resource management must not be an afterthought. But what are resources, and why is it so vital to manage them correctly? A resource is anything you provision: a virtual machine, a database instance, a storage bucket, or anything else your organization creates in GCP.

Projects

To be able to create resources in GCP, it's necessary to create a project. Resources always exist inside projects and cannot be moved to another project after being created, so it is imperative to consider a proper project structure.

While it is possible to have all resources inside a single project, this is a bad idea for several reasons. You'll want employees (and perhaps externals) to be able to create resources and modify existing ones. But you'll also want to restrict them to a set of resources they are allowed to manage. The best approach to access control is to grant employees access to only the projects they need to access to get their work done. Doing so increases security in case of unauthorized access to employee accounts, and, at the same time, decreases the blast radius of any changes.

Also, when executing API operations towards resources in GCP, it is necessary to specify the project in which the API calls are taking effect. A good project structure will ensure that applications and scripts can only affect their relevant resources; even if an API caller had sufficient permissions, it can't affect resources in other projects.

The actual project structure will depend on the organization's complexity and data security requirements.

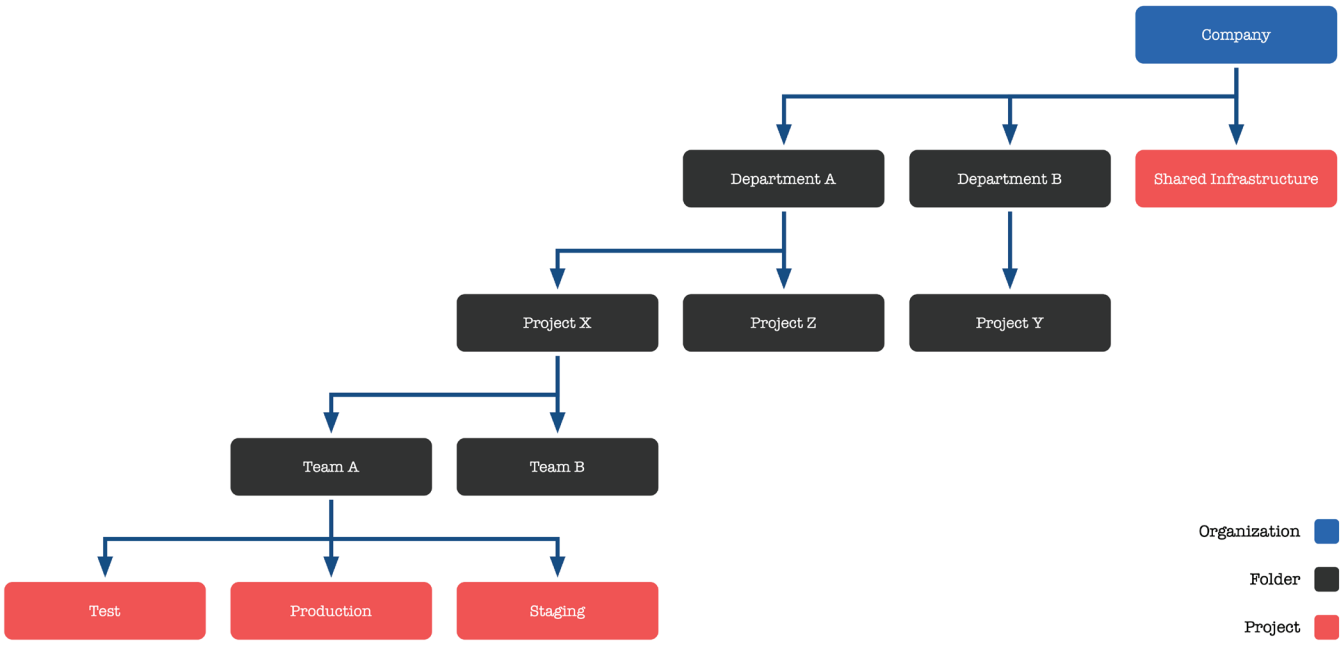
Hierarchies

In addition to projects, there are two other hierarchical resource types: organizations and folders. If your company is already using G Suite or Cloud Identity, then you already have an Organization tied to your company domain name. For users that signed up to Google Cloud using a regular Gmail account, there is no Organization resource automatically created.

While the Organization resource is not required, the owner of a project without an Organization resource in GCP is the employee who created it. Upon employment termination, the ex-employee would have to transfer the project to another GCP user. When using the Organization resource, the company keeps ownership of all of its resources, regardless of which employee created them. For businesses, regardless of size, it is always recommended to use an Organization resource, either by using G Suite, or through Cloud Identity, which can synchronize with an existing Active Directory or LDAP directory.

Folders are the last of the management resources available. They exist between organizations and projects. Folders can be nested, just like folders on your computer. A significant advantage of using folders is that IAM policies are inherited, and folders can have IAM policies the same way that Organization resources and projects can. By using policy inheritance correctly, it can become much easier to manage and secure access to resources.

One final advantage of hierarchies is cost management. The billing portal in GCP shows a cost breakdown per folder and project, allowing you to keep different organizational units and team expenses in check.



Example hierarchy between organization, folders, and projects

Constraints

One last aspect of resource management that's worth considering is organization policies. Although the name implies that these policies only apply to Organization resources, it is possible to enforce constraints through these policies in projects and folders as well. And just like IAM policies, they are also inherited.

Constraints do not define who can access which resources (IAM does); instead, they define which resources are allowed. These policies can help ensure compliance. They can, for example, restrict GCP services that are not permitted by the organization and limit the number of geographical regions to an allowed list.

Best Practices

Create multiple small projects if at all possible, especially in larger organizations. With small projects, in the case of a buggy batch change to resources, the maximum it can affect is a single project. At minimum, separate your different environments (e.g., test, staging, production) into separate projects so that changes to resources in a test environment can never affect production resources, even in the case of a faulty script. This separation can also restrict employees from accessing production environments and their data when they have no reason to do so, while still allowing them to use the development environments they need for everyday work.

Use new projects for experiments. When you finish an experiment, deleting a project deletes all of its resources automatically. By enforcing this practice throughout your organization, employees can quickly try out new ideas and GCP products without the risk of making changes to unrelated resources. There is also no need to worry about forgetting to delete potentially expensive resources when the experimentation phase is over.

Cloud IAM

Identity and Access Management (IAM) is often daunting to someone who hasn't encountered it before. However, there are just a few key concepts to understand. The purpose of Cloud IAM is to control access to cloud resources.

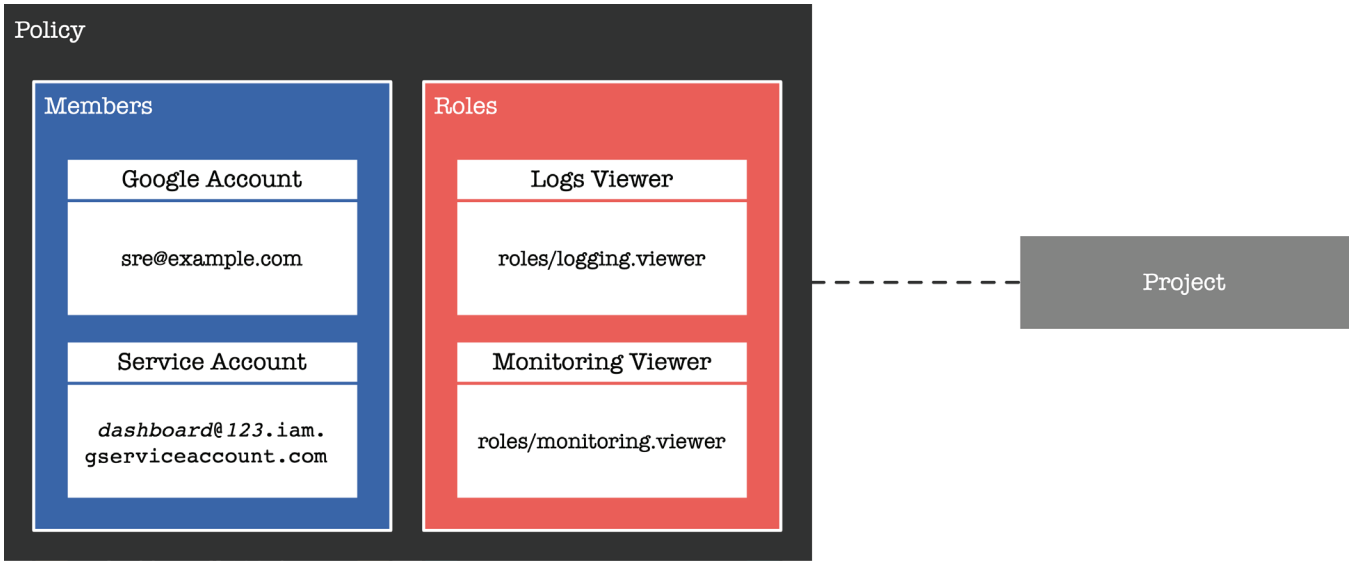
Members, Roles, and Policies

Members define who has access. A member can be anyone who can log in through a Google Account, which can be either a personal account or a G Suite or Google Cloud Identity user.

Because managing individual accounts separately for IAM can be a repetitive task, it's also possible to assign an entire Google Group or domain from G Suite or Google Cloud Identity. A member may also be software that needs to access GCP resources, either running inside or outside of GCP. Service accounts (described below), exist for precisely this purpose.

Roles define a collection of permissions that are allowed on resources. Permissions themselves are very granular. For example, listing Google Cloud Storage buckets, listing the objects inside buckets, and reading the objects themselves are all separate permissions. But by combining all three permissions into a single role, it can now be reused for anyone who needs this kind of read access. Hundreds of pre-made roles exist for all GCP services. It is also possible to create custom roles from any set of permissions.

Lastly, policies are the glue between members, roles, and resources. A policy binds a set of members to a set of roles and is attached to a resource. As previously mentioned in resource management, IAM policies are inherited. A policy can be attached to a single resource or to a project, folder, or organization, in which case it would apply to all resources under it.



Relationship between policies, members, roles, and projects

Service Accounts

Service accounts are accounts generated for applications and virtual machines and do not belong to users. Service accounts do not have passwords, and cannot log into the Google Cloud Console.

To programmatically authenticate using a service account, a service account key pair is needed. While it is possible to supply Google with just a public key to use, it's usually a better idea to let GCP handle the service account keys so that you don't need to worry about key rotation best practices. Cloud IAM can generate Google-managed keys, or you can use the Cloud Key Management Service (KMS).

Best Practices

Always follow the principle of least privilege. Attach policies to the lowest level resource required. For employees, this may be a folder or project. For applications using service accounts that only need a handful of resources, attach policies directly to the resources if possible.

Take advantage of resource management. A well-thought-out hierarchy with folders and projects makes it easier to apply IAM policies at the correct level, making the whole organization more secure. At the same time, remember that policies are inherited. Fortunately, when seeing the policies for any resource, the Google Cloud Console displays what the attached policies versus inherited ones are.

Networking

So far, various security topics have been discussed: the building blocks of GCP cloud security, resource management, and IAM. This section will discuss connections among all the pieces.

Virtual Private Cloud

Networking in a cloud environment isn't much different from networking in a traditional data center, since all of the same concepts, rules, and best practices apply. One key difference between the processes is that in the cloud, everything is software-defined. You no longer have to worry about the physical limitations imposed by networking hardware.

When using GCP, you get a private, fully managed network in the form of a Virtual Private Cloud (VPC). This VPC is isolated from every other network in Google Cloud.

GCP's VPC network is unique because it is global. Some other cloud providers offer VPC networks that exist in a single region and require peering connections between VPCs in different regions. GCP's network not only makes managing a globally-distributed network much more straightforward (i.e., there is no need for peering connections or VPNs), it also keeps traffic between regions off the public internet.

VPC Concepts

For a network to be usable, there must be at least one subnet, which is a subset of the total IP address space of the network. While the network itself is global, the subnet is regional. Resources that involve networking, such as compute instances, are always attached to a subnet.

Routes determine what happens to traffic that leaves a subnet, commonly referred to as egress traffic. After you create a new subnet, a subnet route is automatically established by the system, making it possible to communicate between subnets. The system also automatically creates a default route that enables internet access. It is possible to delete this default route in order to isolate the network from the public internet. You can also replace the route completely and customize the egress traffic's path to the internet.

Firewall rules control ingress and egress traffic. These rules not only control the traffic within the network, they also control traffic coming into and leaving the network. A firewall rule includes the

direction of traffic (ingress or egress), the priority of the rule represented by a number from 0 (highest) to 65535 (lowest), the action to take (allow or deny), a target, and a source or destination (depending on the direction of traffic).

Best Practices

Simpler network topologies are much easier to understand, and a better understanding of the overall network results in a lower chance of hidden security holes. To make the network easier to manage and understand, you should group VMs with similar network security requirements in the same subnet and apply firewall rules to the entire subnet.

For production, always create a well-planned custom network. By default, new GCP projects contain an automatically-created network with a subnet for each region. All default networks have the same IP address range, which creates an overlap when connecting to other default networks. For existing projects, the default network can be deleted at any time, provided it's not in use. It's also possible to disable the creation of default networks for new projects by using an organization policy with a constraint.

Limiting external access is one of the simplest forms of security, and it is still a highly effective one. When external IP addresses on VMs are disabled, it is impossible to initiate connections to them from the internet. If you still want to allow these instances to be able to connect to the internet (for external API access or updates), you can use a NAT instance.

Just as you should keep different environments in different projects, you should also **keep different environments in different networks.**

Gotchas

One important limitation to be aware of with VPC is that only IPv4 is supported. While public internet traffic from and to a GCP load balancer can use IPv6, it is converted into IPv4 traffic before entering the VPC.

There are two implied firewall rules that don't show up in the Google Cloud Console. Although you cannot delete them, they exist at the lowest possible priority, making them easy to override. One rule denies all ingress traffic with a source 0.0.0.0/0 (i.e., from anywhere)—a good default. The other allows most egress traffic with a destination of 0.0.0.0/0, which may or may not be what you want. It's imperative to be aware of these implied rules and override them to suit your security requirements.

Network Monitoring

Monitoring is crucial not only for troubleshooting but also for security. Fortunately, monitoring one or more VPCs is quite straightforward.

VPC traffic can be logged as it enters and leaves compute instances. These logs are called VPC Flow Logs. They record information about the TCP and UDP traffic and enable you to monitor the performance and throughput of your network, helping you to better plan your capacity. They can also be used to detect unwanted/hostile traffic or assist in debugging efforts.

You can adjust the logs' sampling levels from 0% to 100% of traffic. The aggregation time interval, which defaults to 5 seconds, can also be adjusted to give you near real-time log processing capabilities.

Enabling flow logs per subnet is also possible. Because flow logs use Stackdriver Logging, you can add filters to them and even forward them to other services supported by Stackdriver, such as Cloud Pub/Sub for real-time processing and BigQuery for long-term storage.

It is worth understanding how the VPC Flow Logs are generated. Each log entry is made up of aggregated ingress and egress packets on VMs. However, only about 1 in 10 packets are sampled. While this may not be a problem, it needs to be taken into account when planning your security and compliance requirements.

You can also monitor and audit firewall rules using Firewall Rules Logging. This feature can be incredibly helpful in determining whether or not a firewall rule is acting as expected. It can also monitor the amount of traffic that is allowed and denied. For each firewall rule, the logs can quickly be enabled or disabled. As with flow logs, firewall rules logs are sent to Stackdriver Logging, allowing you to take advantage of all the same advanced search, filtering, and exporting capabilities.

Best Practices

It is worth the effort to **use these capabilities and monitor traffic as part of your daily workflows**, whether via the native GCP products or by connecting them into your SIEM/SOC. Even better is to **integrate them with your web security solution(s)**.

A common security mistake is to only pay attention to traffic that has been blocked; indeed, many security solutions only report the requests that were identified as hostile. However, when diagnosing a traffic issue, it is frequently valuable to examine all incoming requests (passed and blocked), and their disposition. Comparing blocked requests to those that are passed allows you to understand exactly

how traffic is being processed, and can help diagnose traffic issues that otherwise could be difficult to understand.

A few security solutions provide this capability out of the box. Reblaze displays all incoming requests and their disposition in real time, and also streams its data into Google Cloud Security Command Center. It stores all traffic data in BigQuery, thus providing comprehensive historical logs.

Network Address Translation

As previously mentioned, it's often necessary to prevent virtual machines from accessing the internet directly—either for security or compliance reasons, or to have a static list of public IPs that your partners and clients can use to whitelist your traffic.

Instead of having to manage additional virtual machines to perform network address translation, GCP provides Cloud NAT, a managed, scalable, high-performance service that addresses this exact need. And by using a managed service, your share of responsibility decreases. You don't have to harden and secure VMs running at the edge of your network or shoulder a maintenance and operational burden.

Please note that a single Cloud NAT can have multiple public IP addresses and subnets, but it belongs to a single region. For networks with multiple regions requiring NAT, at least one Cloud NAT must be created per region.

Connecting Networks

Eventually, you'll need to connect two or more networks. Whether the networks are all in GCP or include external networks, such as an on-premises network or another cloud provider, Google Cloud provides a plethora of possibilities for linking them.

Peering and Sharing

Connecting networks inside Google Cloud is a straightforward process known as VPC peering. It is possible to create a peering connection between any two GCP networks, even if they reside in different organizations. Compared to VPN solutions, peering is much simpler, and it is available at no extra cost.

However, for organizations that have multiple projects with shared networking, peering is not the only solution. It is possible to create a separate GCP project (a host project) with a shared VPC network.

Other projects (service projects) are then associated with the shared network so that they can launch resources either in the entire shared network, which is the default, or just in one or more subnets, which is achieved by creating an organization policy constraint.

If you're starting to plan your network from scratch, check to see if a shared network can suit your needs since VPC peering comes with limitations (such as the number of maximum peering connections) that are not present in a shared VPC network.

Virtual Private Network (VPN)

When it's necessary to connect to networks outside of Google Cloud, you can use Cloud VPN, a fully managed, low-cost VPN solution. In it, each single Cloud VPN tunnel supports up to 3Gbps of network bandwidth. Cloud VPN supports high availability (HA) setups with multiple VPN tunnels, in which case an SLA of 99.99% takes effect. IPSec as well as both IKEv1 and IKEv2 are supported by it, as is dynamic (BGP) routing.

Cloud Interconnect

If a physical connection to Google Cloud is required, GCP offers Cloud Interconnect. Because Cloud Interconnect is a physical connection outside of the public internet, it can be more cost-effective for large volumes of transferred data. There are two types of interconnect options. The first is a dedicated pipe from your network to one of Google's point-of-presence (POP) locations. This pipe provides 10Gbps and 100Gbps circuits, up to a maximum of 200Gbps per connection. The second interconnect option is necessary when connecting directly to a Google POP is not feasible, or when less network throughput is required. In the latter case, there is a long list of partners that provide between 50Mbps to 50Gbps of connectivity. The partners maintain most of the required hardware.

Unlike VPN connections, direct connections to Google Cloud provide no encryption of the traffic whatsoever. This means that encryption has to be handled at the application level.

Best Practices

When designing network topologies, any network that is accessible from the public web must have robust protection against Internet threats: hacking and breach attempts, DDoS attacks, hostile bots, and so on. A comprehensive security solution such as Reblaze (discussed on pages 6-7) can scrub incoming web traffic.

Conclusion

This white paper has provided a high-level overview of the issues involved when securely using Google Cloud Platform.

As mentioned in the introduction, for more in-depth information on GCP topics it's worth the time to browse the GCP YouTube channel. For an overview of the various services available within GCP, the Products page is the best place to start.

Securely using GCP is a front-loaded process, with much of the work done in the beginning. Much of the necessary setup and configuration is fire-and-forget. However, there is one area for which this is not true: **web security**. Blocking attack traffic that is attempting to access your applications, services, and APIs is an ongoing process, evolving constantly as the threats themselves grow and change.

Reblaze is a fully managed cloud platform, providing comprehensive, effective web security for GCP, automating its inherent security capabilities and adding many more. Machine Learning provides accurate, adaptive threat detection; dedicated Virtual Private Clouds ensure maximum privacy, performance, and protection. For more information, or to get a demo, contact us here.

Questions about the content of this white paper? Contact us at hello@reblaze.com.

To receive notifications of future publications, sign up for our newsletter by filling out the form at www.reblaze.com/contact-us/.

Reblaze Technologies, Ltd.
3031 Tisch Way
110 Plaza West
San Jose, CA 95128