Infraestructura de Google

Estrella Afán de Rivera Díaz

Javier Oliva Cruz

Fernando Calvillo Parejo

Índice de contenido

1. Introducción.

La infraestructura de Google está diseñada para brindar seguridad durante todo el ciclo de vida del procesamiento de la información.

Esta infraestructura proporciona implementación segura de servicios, almacenamiento seguro de datos, comunicaciones seguras entre servicios, comunicación segura y privada con los clientes a través de Internet y operaciones seguras por parte de los administradores.

Google utiliza esta infraestructura para construir sus servicios de Internet, incluidos servicios de consumo como Búsqueda, Gmail y Fotos, y servicios empresariales como G Suite y Google Cloud Platform.

Primero se mencionarán las capas que forman dicha infraestructura, para después describir con más detalle cómo desarrolla Google la seguridad de todas estas capas.

2. Capas de seguridad de la infraestructura de Google.

En la siguiente tabla se representan las diversas capas de seguridad, desde la infraestructura de hardware en la capa inferior hasta la seguridad operacional en la capa superior, seguidas por los mecanismos que usa Google para garantizar la seguridad de cada parte.

- Seguridad operacional -			
Detección de intrusión	Reducir riesgo interno	Dispositivos y credenciales de empleados seguros	Desarrollo de software seguro
- Comunicación de Internet -			
Front-end de Google	Protección DoS		
- Servicios de almacenamiento -			
Cifrado en reposo	Eliminación de datos		
- Identidad del usuario -			
Autenticación	Protección de abuso de login		
- Despliegue del servicio -			
Gestión de acceso de datos de usuario final	Cifrado de la comunicación entre servicios	Gestión de acceso entre servicios	Identidad de servicio, integridad y aislamiento
- Infraestructura hardware -			
Pila de arranque e identidad de la máquina seguras	Diseño y procedencia del hardware	Seguridad de las instalaciones físicas	

A continuación se describe detalladamente el contenido de cada capa, comenzando con la capa de más bajo nivel.

3. Seguridad en la infraestructura hardware.

En esta sección se describe cómo Google protege las capas más bajas de su infraestructura, desde las instalaciones físicas hasta el hardware especialmente diseñado en sus centros de datos, así como la pila de software de bajo nivel que se ejecuta en cada máquina.

1. Seguridad de las instalaciones físicas.

Google diseña y construye sus propios centros de datos, que incorporan múltiples capas de protecciones de seguridad física. El acceso a estos centros de datos está limitado a sólo una pequeña fracción de los empleados de Google.

Utilizan múltiples capas de seguridad física para proteger los pisos de sus centros de datos, y también utilizan tecnologías como identificación biométrica, detección de metales, cámaras, barreras para vehículos y sistemas de detección de intrusos basados en láser.

Además, Google aloja algunos servidores en centros de datos de terceros, donde se aseguran de que haya medidas de seguridad física controladas por Google por encima de las capas de seguridad proporcionadas por el operador del centro de datos.

2. Diseño y procedencia del hardware.

Un centro de datos de Google consta de miles de máquinas de servidor conectadas a una red local. Tanto las placas de servidor como el equipo de red están diseñados a medida por Google. Revisan a los proveedores de componentes con los que trabajan y eligen los componentes con cuidado. También diseñan chips personalizados, que permiten identificar y autenticar de forma segura los dispositivos legítimos de Google a nivel de hardware.

3. Asegurar la pila de arranque y la identidad de la máquina.

Las máquinas servidor de Google utilizan una variedad de tecnologías para garantizar que están arrancando la pila de software correcta. Utilizan firmas criptográficas en componentes de bajo nivel como el BIOS, el cargador de arranque, el kernel y la imagen del sistema operativo base. Estas firmas se pueden validar durante cada arranque o actualización. Los componentes son todos controlados, construidos y endurecidos por Google.

Cada máquina servidor en el centro de datos tiene su propia identidad específica, la cual se utiliza para autenticar llamadas de API hacia y desde servicios de administración de bajo nivel en la máquina.

Google ha creado sistemas automatizados para garantizar que los servidores ejecuten versiones actualizadas de sus pilas de software (incluidos los parches de seguridad), para detectar y diagnosticar problemas de hardware y software, y eliminar las máquinas del servicio si es necesario.

4. Seguridad en el despliegue del servicio.

En este punto se describirá como Google pasa del hardware y software para garantizar que un servicio se implementa de forma segura en sus infraestructuras.

La infraestructura no asume ninguna confianza entre los servicios que se ejecutan en ella. En otras palabras, la infraestructura está diseñada fundamentalmente para ser multi-arrendataria.

1. Identidad, integridad y aislamiento del servicio.

   Google utiliza la autenticación y autorización criptográfica en la capa de aplicación para la comunicación entre servicios. Esto proporciona un control de acceso sólido a un nivel de abstracción y granularidad que los administradores y servicios pueden comprender de forma natural.

   Cada servicio que se ejecuta en la infraestructura tiene una identidad de cuenta de servicio asociada. A un servicio se le proporcionan credenciales criptográficas que puede usar para probar su identidad al realizar o recibir llamadas a procedimientos remotos (RPC) a otros servicios. Los clientes utilizan estas identidades para asegurarse de que están hablando con el servidor correcto y los servidores para limitar el acceso a métodos y datos a clientes particulares.

   El código fuente de Google se almacena en un repositorio central donde las versiones actuales y pasadas del servicio son auditables. La infraestructura puede configurarse adicionalmente para requerir que los binarios de un servicio se construyan a partir de un código fuente específico revisado, registrado y probado. Dichas revisiones de código requieren la inspección y aprobación de al menos un ingeniero que no sea el autor, y el sistema hace cumplir que las modificaciones del código a cualquier sistema deben ser aprobadas por los propietarios de ese sistema. Estos requisitos limitan la capacidad de un interno o adversario para realizar modificaciones maliciosas en el código fuente y también proporcionar un rastro forense desde un servicio hasta su fuente.

   Google tiene una variedad de técnicas de aislamiento y sandbox para proteger un servicio de otros servicios que se ejecutan en la misma máquina. Estas técnicas incluyen la separación normal de usuarios de Linux, entornos limitados basados en el lenguaje y el kernel y la virtualización de hardware.

2. Gestión de acceso entre servicios.

   El propietario de un servicio puede usar las funciones de administración de acceso proporcionadas por la infraestructura para especificar exactamente qué otros servicios pueden comunicarse con él. Ese servicio se puede configurar con la lista blanca de las identidades de cuenta de servicio permitidas y esta restricción de acceso se aplica automáticamente por la infraestructura.

   Los ingenieros de Google que acceden a los servicios también reciben identidades individuales, por lo que los servicios pueden configurarse de manera similar para permitir o denegar sus accesos. Todos estos tipos de identidades (máquina, servicio y empleado) están en un espacio de nombres global que mantiene la infraestructura.

   La infraestructura proporciona un sistema de flujo de trabajo de gestión de identidades para estas identidades internas, incluidas las cadenas de aprobación, el registro y la notificación. Por ejemplo, estas identidades se pueden asignar a grupos de control de acceso a través de un sistema que permite el control de dos partes, donde un ingeniero puede proponer un cambio a un grupo que otro ingeniero (que también es administrador

del grupo) debe aprobar. Este sistema permite que los procesos de administración de acceso seguro se amplíen a los miles de servicios que se ejecutan en la infraestructura.

3. Cifrado de la comunicación entre servicios.

   La infraestructura también proporciona privacidad e integridad criptográfica para los datos de RPC en la red. Para proporcionar estos beneficios de seguridad a otros protocolos de la capa de aplicación como HTTP, se encapsulan dentro de los mecanismos de RPC de la infraestructura. En esencia, esto proporciona un aislamiento de la capa de aplicación y elimina cualquier dependencia de seguridad de la ruta de red. La comunicación cifrada entre servicios puede permanecer segura incluso si se toca la red o si se compromete un dispositivo de red.

   Los servicios pueden configurar el nivel de protección criptográfica que desean para cada RPC de la infraestructura. Para protegerse contra los que pueden estar intentando acceder a los enlaces WAN privados, la infraestructura encripta automáticamente todo el tráfico RPC de infraestructura que pasa por la WAN entre centros de datos, sin requerir ninguna configuración explícita del servicio.

4. Gestión de acceso de datos de usuario final.

   Hemos visto en la anteriormente que el servicio de Contactos se puede configurar de manera que las únicas solicitudes de RPC que se permiten sean las del Servicio de Gmail.

   Esto, sin embargo, sigue siendo un conjunto muy amplio de permisos. Dentro del alcance de este permiso, el servicio de Gmail podría solicitar los contactos de cualquier usuario en cualquier momento.

   Dado que el servicio de Gmail realiza una solicitud de RPC al servicio de Contactos en nombre de un usuario final en particular, la infraestructura proporciona una capacidad para que el servicio de Gmail presente un "ticket de permiso de usuario final" como parte del RPC. Este ticket prueba que el servicio de Gmail está atendiendo actualmente una solicitud en nombre de ese usuario final en particular. Esto permite que el servicio de contactos implemente una salvaguardia donde solo devuelve datos para el usuario final nombrado en el ticket.

   La infraestructura proporciona un servicio de identidad de usuario central que emite estos "tickets de permiso de usuario final". Un inicio de sesión de usuario final es verificado por el servicio central de identidad que luego emite una credencial de usuario. Cada solicitud posterior del dispositivo cliente a Google debe presentar esa credencial de usuario.

   Cuando un servicio recibe una credencial de usuario final, pasa la credencial al servicio central de identidad para su verificación. Si la credencial del usuario final se verifica correctamente, el servicio de identidad central devuelve un "ticket de permiso de usuario final" de corta duración que se puede usar para los RPC relacionados con la solicitud.

5. Almacenamiento seguro de datos.

Hasta ahora se ha descrito como Google implementa los servicios de forma segura. En este punto se explicará cómo implementa el almacenamiento seguro de datos en la infraestructura.

1. Cifrado en reposo.

   La infraestructura de Google aporta diversidad de servicios de almacenamiento, como por ejemplo BigTable y Spanner, y un servicio central de administración de claves. La mayoría de las aplicaciones de Google acceden de forma indirecta a través de estos servicios de almacenamiento al almacenamiento físico. Este servicio de administración de claves admite la rotación automática de claves, proporciona extensos registros de auditoría y se integra con los tickets de permiso del usuario final mencionados anteriormente para vincular claves a usuarios finales particulares.

   En la capa de aplicación la ejecución del cifrado permite a la infraestructura aislarse de posibles amenazas en los niveles más bajos de almacenamiento, como el firmware de disco malicioso. Además, también hay capas adicionales de protección en la infraestructura. Google habilita el soporte de cifrado de hardware en en sus HDD y SSD y rastrean cada unidad a través de su ciclo de vida. Antes de que un dispositivo de almacenamiento cifrado retirado del servicio pueda abandonar físicamente las instalaciones de Google, se limpia mediante un proceso de varios pasos que incluye dos verificaciones independientes. Los dispositivos que no pasan este procedimiento de limpieza son destruidos físicamente (por ejemplo, triturados) en las instalaciones.

2. Eliminación de datos.

   La eliminación de datos en Google a menudo comienza con la marcación de datos específicos como "programados para eliminación" en lugar de eliminar los datos por completo. Esto les permite recuperarse de eliminaciones no intencionales, ya sean iniciadas por el cliente o debido a un error o error de proceso interno. Después de haber sido marcado como "programado para su eliminación", los datos se eliminan de acuerdo con las políticas específicas del servicio.

6. Comunicación segura de internet.

Hasta este punto, se ha descrito como Google asegura los servicios en su infraestructura. En esta sección, se va a describir cómo protege la comunicación entre Internet y estos servicios.

Como se mencionó anteriormente, la infraestructura consiste en un gran conjunto de máquinas físicas que están interconectadas a través de LAN y WAN, y la seguridad de la comunicación entre servicios no depende de la seguridad de la red. Sin embargo, Google aísla su infraestructura de Internet en un espacio de IP privado para que se puedan implementar protecciones adicionales más fácilmente.

1. Servicio de front-end de Google.

   Cuando un servicio quiere estar disponible en Internet, puede registrarse con un servicio llamado Google Front End (GFE). El GFE garantiza que todas las conexiones TLS se terminen utilizando certificados correctos y siguiendo las mejores prácticas. El GFE además aplica protecciones contra los ataques de Denegación de Servicio. El GFE luego

reenvía las solicitudes del servicio utilizando el protocolo de seguridad RPC.

2. Protección de denegación de servicio (DoS).

   La gran escala de su infraestructura permite a Google simplemente absorber muchos ataques DoS. Dicho esto, tienen protecciones DoS de niveles múltiples y capas que reducen aún más el riesgo de cualquier impacto DoS en un servicio que se ejecuta detrás de un GFE.

   Después de que su red troncal proporcione una conexión externa a uno de sus centros de datos, pasa a través de varias capas de equilibrio de carga de hardware y software. Estos equilibradores de carga dan información sobre el tráfico entrante a un servicio DoS central que se ejecuta en la infraestructura. Cuando el servicio DoS central detecta que se está produciendo un ataque DoS, puede configurar los balanceadores de carga para que eliminen o aceleren el tráfico asociado con el ataque.

   En la siguiente capa, las instancias de GFE también reportan información sobre las solicitudes que están recibiendo al servicio DoS central, incluida la información de la capa de aplicación que los balanceadores de carga no tienen. El servicio DoS central también puede configurar las instancias de GFE para que eliminen o aceleren el tráfico de ataque.

3. Autenticación de usuario.

   Después de la protección DoS, la siguiente capa de defensa proviene de su servicio de identidad central. Este servicio generalmente se manifiesta a los usuarios finales como la página de inicio de sesión de Google. Además de solicitar un nombre de usuario y una contraseña simples, el servicio también reta inteligentemente a los usuarios para obtener información adicional basada en factores de riesgo, como si han iniciado sesión desde el mismo dispositivo o una ubicación similar en el pasado. Después de autenticar al usuario, el servicio de identidad emite credenciales como cookies y tokens de OAuth que se pueden usar para llamadas posteriores.

   Los usuarios también tienen la opción de emplear segundos factores como OTP o claves de seguridad resistentes al phishing al iniciar sesión.

7. Seguridad operacional.

A continuación, se va a describir cómo operar en la infraestructura de manera segura. Para ello se creará un software de infraestructura seguro, se protegerán las máquinas y credenciales de los empleados y se defenderán las amenazas a la infraestructura tanto del personal interno como de factores externos.

1. Desarrollo de software seguro.

   Se ofrecen bibliotecas que impiden que los desarrolladores puedan introducir ciertos errores de seguridad. Por ejemplo, hay bibliotecas que eliminan las vulnerabilidades XSS en la aplicación web. Además, google cuenta con herramientas automatizadas para detectar automáticamente errores de seguridad incluidos en fuzzers, herramientas de análisis estático y escáneres de seguridad web.

   Aparte de lo explicado, se realizan verificaciones finales de seguridad manual, como revisiones de implementación y diseño en profundidad para las funciones de mayor

riesgo.

Estas revisiones son realizadas por un equipo que incluye expertos en seguridad web, criptografía y seguridad del sistema operativo. Las revisiones también pueden dar como resultado nuevas características de la biblioteca de seguridad y nuevos fuzzers que luego se pueden aplicar a otros productos futuros.

Además, existe un Programa de recompensas de vulnerabilidad en el que Google paga a cualquier persona que pueda descubrir e informar sobre errores en la infraestructura o aplicaciones.

2. Mantener seguros los dispositivos y credenciales de los empleados.

Para mantener seguros los dispositivos y credenciales de los empleados se realiza una gran inversión, además se monitorean las actividades internas en busca de actividades ilícitas.
Se realiza una gran inversión en la supervisión de los dispositivos que los empleados utilizan para operar en la infraestructura de la empresa. Google se asegura de que las imágenes del sistema operativo para estos dispositivos estén actualizadas con parches de seguridad y se controlan las aplicaciones que se pueden instalar. Además, contamos con sistemas para escanear aplicaciones instaladas por el usuario, descargas, extensiones de navegador y contenido explorado desde la web para verificar su idoneidad en los clientes corporativos.

8. Conclusión.

Se ha descrito cómo la infraestructura de Google está diseñada para construir, implementar y operar servicios de manera segura a escala de Internet. Esto incluye servicios al consumidor como Gmail y sus servicios empresariales. Además, las ofertas de Google Cloud están construidas sobre esta misma infraestructura.

Google invierte fuertemente en asegurar su infraestructura. Tienen cientos de ingenieros dedicados a la seguridad y la privacidad distribuidos en todo Google, incluidos muchos que son autoridades reconocidas de la industria.