Google Cloud

How Google Uses Encryption to Protect Your Data

G Suite Encryption Whitepaper

Table of Contents

Introduction

Here at Google, we know that security is a key consideration for organizations that choose G Suite. This is why we work so hard to protect your data — whether it's traveling over the Internet, moving between our data centers or stored on our servers.

A central part of our comprehensive security strategy is encryption technology, which helps prevent information from being accessed in the event that it falls into the wrong hands. This paper will describe Google's approach to encryption and how it keeps your sensitive information safe.

How Google Approaches Encryption

Encryption works by replacing data with unreadable code known as ciphertext. To decrypt the ciphertext back into its original form, you need to employ the key used in the encryption algorithm. Attackers who want to circumvent encryption will typically try to steal the keys or exploit flaws in the encryption algorithms and their implementation. Encryption strength depends on a number of factors, such as how keys are created, managed and secured. It also depends on the algorithm used and the key size for that algorithm. As computers get better and faster, it becomes easier to perform the complicated mathematical computations needed to break encryption. Even the mathematics behind this process — known as cryptanalysis — can improve over time, making it easier to break encryption. As a result, encryption algorithms that seemed strong a few years ago may no longer be as strong today.

To keep pace with this evolution, Google has a team of world-class security engineers tasked with following, developing and improving encryption technology. Our engineers take part in standardization processes and in maintaining widely used encryption software such as OpenSSL. We regularly publish our research in the field of encryption so that everyone in the industry — including the general public — can benefit from our knowledge. For example, in 2014 we revealed a significant vulnerability in SSL 3.0 encryption known as POODLE and in July of 2015 a high-level vulnerability in OpenSSL.

Google has a team of world-class security engineers tasked with following, developing and improving encryption technology.

Encryption is an important piece of the G Suite security strategy, helping to protect your emails, chats, Google Drive files and other data. First, we encrypt certain data as described below while it is stored “at rest” — stored on a disk (including solid-state drives) or backup media. Even if an attacker or someone with physical access obtains the storage equipment containing your data, they won’t be able to read it because they don’t have the necessary encryption keys. Second, we encrypt all data while it is “in transit” — traveling over the Internet and across the Google network between data centers. Should an attacker intercept such transmissions, they will only be able to capture encrypted data. We’ll take a detailed look at how we encrypt data stored at rest and data in transit below.

Encryption of Data Stored at Rest

Data belonging to G Suite customers is stored at rest in two types of systems: disks and backup media. Disks are used to write new data as well as store and retrieve data in multiple replicated copies. (For more information on this replication of data, please see the G Suite Security Whitepaper.) Google also stores data on offline backup media to help ensure recovery from any catastrophic error or natural disaster at one of our data centers. Data stored at rest is encrypted on both disks and backup media, but for each system we use a distinct approach for encryption to mitigate the corresponding security risks. These encryption mechanisms are detailed below.

Data on disks

Google encrypts customers' data stored at rest for the solutions in the G Suite product family (see Table 1). This encryption happens without the customer having to take any action. Core content is data created by the user, such as messages and attachments in Gmail.

To understand how this encryption works, it's important to understand how Google stores customer data. Data is broken into subfile "chunks," which are stored on local disks and identified by unique chunk IDs.

Google encrypts data as it is written to disk with a per-chunk encryption key that is associated with a specific Access Control List (ACL). The ACL ensures that data in each chunk is only decrypted by authorized Google employees and services that were given permission at the time of encrypting the data.

Google encrypts data as it is written to disk with a per-chunk encryption key that is associated with a specific Access Control List.

This means that different chunks are encrypted with different encryption keys, even if they belong to the same customer. These chunks are encrypted using 128-bit or stronger Advanced Encryption Standard (AES).

Table 1 details what type of data is encrypted by each G Suite solution.

Table 1

G Suite: Encryption of data stored at rest

Solution	Core Content Data that Is Encrypted
Gmail	Messages and attachments
Calendar	Events and descriptions of events
Drive	Files uploaded to Drive via Google Drive for Windows and Mac, via the Drive web interfaces, Drive Mobile apps, Google Drive API, Google Photos, and Gmail. In all these cases videos uploaded may not be encrypted.
Docs	Content authored by the owner or collaborators of the doc, except content embedded into the doc that is hosted on other Google products not referenced in this list (e.g., YouTube)
Sheets (including Forms)	Content authored by the owner or collaborators of the spreadsheets, except content embedded into the spreadsheets that is hosted on other Google products not referenced in this list (e.g., YouTube)
Slides	Content authored by the owner or collaborators of the presentation, except content embedded in the presentation that is hosted on other Google products not referenced in this list (e.g., YouTube)
Talk	Archived "on the record" conversations
Hangouts chat	Archived "on the record" conversations[1]
Sites	Content authored by the owners or collaborators of the site; except (i) content embedded into the site that is hosted on other Google products not referenced in this list (e.g., YouTube) (ii) content embedded into the site that remains hosted on other third-party websites, via Sites, Gadgets or image hotlinking
Contacts	Content of end users' address books
Groups	Group message archives
Vault	Content created by Vault Admins is encrypted; saved queries, audit logs are encrypted. Vault's exports of Gmail, messages and attachments, Talk conversations, Hangouts chat and Drive files (except video content) are also encrypted

[1]Off the record chats are not kept, hence encryption at rest is not applicable

Key management and the decryption process

Managing keys safely and reliably, while allowing access to the keys only to authorized services and individuals, is central to encrypted data security. Google has built a robust proprietary service for the distribution, generation, rotation and management of cryptographic keys using industry standard cryptographic algorithms that are in alignment with stronger industry practices. In the following sections, we'll outline our approach to managing the encryption keys used to protect G Suite customers' information.

Google's key management service

As described in the previous section, files or data structures with customer-created content written by G Suite are subdivided into chunks, each of which is encrypted with its own chunk data encryption key ("chunk key"). Each chunk key is encrypted by another key known as the wrapping key, which is managed by a Google-wide key management service (KMS). The result is a "wrapped" (encrypted) chunk key, which is stored alongside the encrypted data. The wrapping keys, needed to decrypt wrapped chunk keys, and therefore to decrypt the chunk, are known only to the KMS and are never stored at rest in unencrypted form.

Decryption and encryption operations on chunk keys are performed within the KMS. The wrapped chunk key is sent by a storage system to the KMS as a request to be unwrapped (decrypted) in order to access the encrypted data. The KMS authenticates the requesting system and checks the request against both system-level and per-wrapped-key ACLs.

Customer-created content written by G Suite is subdivided into chunks, each of which is encrypted with its own chunk data encryption key.

If this request is authorized, the chunk data key is decrypted in the KMS and returned to the storage system, which can now use that chunk key to decrypt that specific chunk of data. These chunk keys are encrypted in transit, as described below. This process is repeated until all the chunks that compose a specific file or data structure are decrypted, making the data available to the requesting application.

Data cannot be decrypted without both the wrapping key and the wrapped chunk key. Decrypting data therefore requires the cooperation of the storage system (which holds the encrypted data and wrapped chunk key) and the KMS (which holds the wrapping key). The KMS wrapping keys that encrypt the chunk keys are 128-bit or stronger AES keys.

All access to the KMS is controlled by ACLs. Access is restricted to a limited number of individuals and specific applications that require access. Individuals are only provided access after demonstrating a recorded need. Access requests to the KMS by employees is logged for auditing.

Rotating keys to limit risk

Google has built a proprietary system to manage key rotation. Chunk encryption keys and wrapping keys are rotated or replaced regularly, so that if a key were compromised it wouldn't remain useful for decrypting new data indefinitely. When a new chunk is written, it is encrypted with a newly generated chunk encryption key, wrapped with the current version of the appropriate wrapping key. Wrapping keys are typically rotated at least every 90 days. This process reduces data exposure in the event of a key compromise or cryptanalytic attack by limiting the time window in which any given wrapping key or chunk encryption key is used.

The key management server

The KMS, like other Google production services, runs on custom, purpose-built servers that we design and manufacture ourselves. These servers run a custom-designed operating system based on a stripped-down and hardened version of Linux, and are designed for the sole purpose of providing Google services.

Google servers use a homogeneous environment that is maintained by proprietary software that continually monitors systems for binary modifications to ensure that only approved Google software is installed and running on Google servers. The KMS server has the same proprietary software installed on it monitoring for any unapproved modification. If a modification is found on the KMS server that differs from a standard Google image, the server is automatically returned to its official standard image state. These automated, self-healing mechanisms are designed to enable Google to monitor and remediate destabilizing events, receive notifications about incidents, and slow down potential compromise on the network or the KMS server. For more information about Google's custom-built servers and how they are managed, please see the G Suite Security Whitepaper.

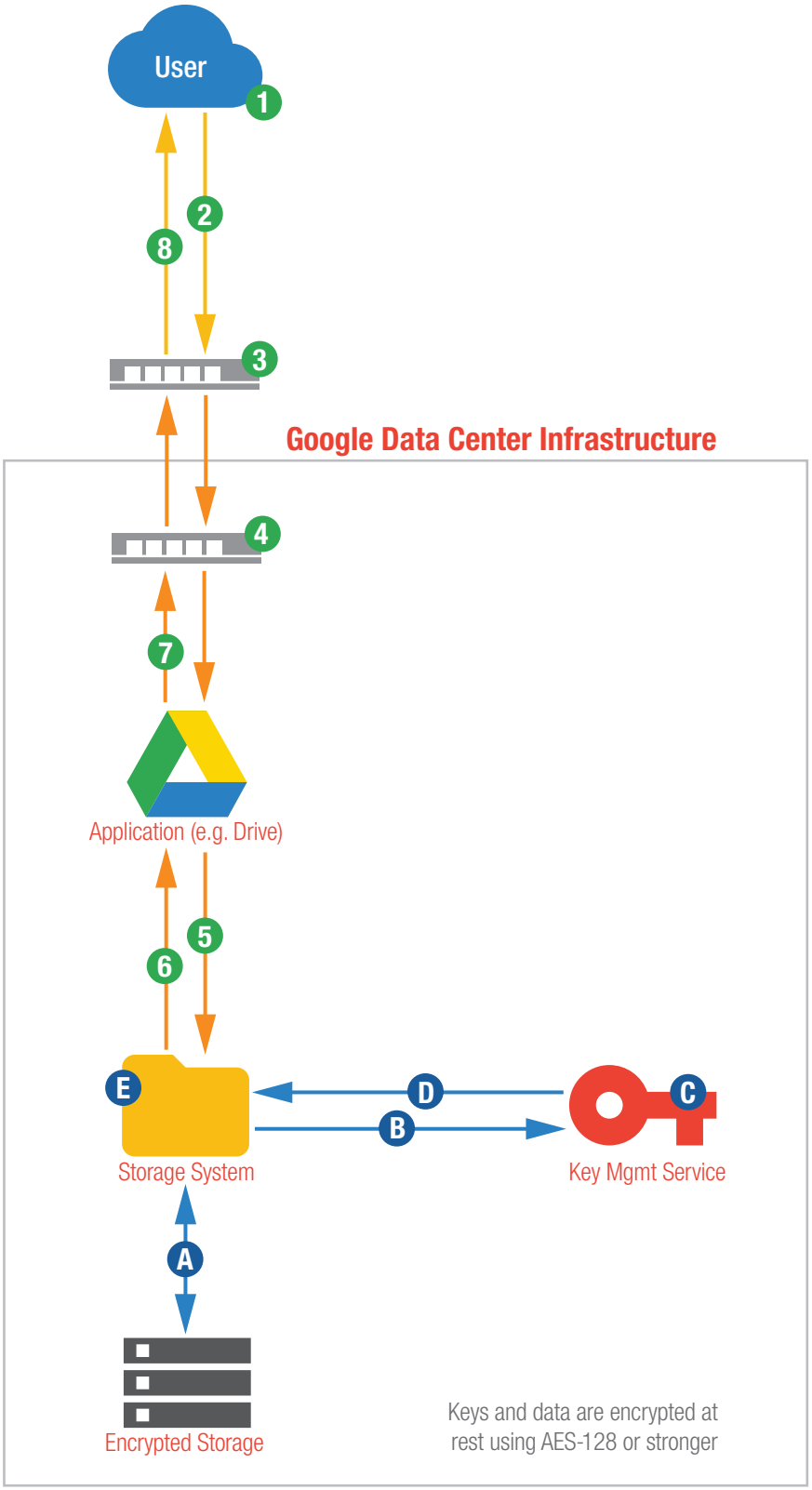
Google has built a system to manage key rotation. Chunk encryption keys and wrapping keys are rotated or replaced regularly.

The following diagram shows the flow for Google Drive, as an example of the related encryption mechanisms. The process begins with the user requesting access to some of their Google Drive data (step 1). The connection between the user and Google is encrypted (step 2). Google routes this request internally (steps 3, 4, 5). When the storage system needs access to an encrypted chunk, Google begins the decryption process (steps A–E) to decrypt the data the user has requested and make it available to Google servers only in memory (i.e., not stored at rest in plaintext). Finally, it returns the data to the user (steps 6, 7, 8), again in an encrypted session. The data flow diagrams are similar for other G Suite products as well as for encrypting data when users create data.

Google Cloud

Encryption at Rest Flow

An example of encryption in Google Drive



USER DATA FLOW

1 Initiate Request
User authenticates to G Suite and requests Drive data.

2 Encrypted Tunnel
SSL/TLS-based encryption dependent on user's browser capabilities.

3 Google Front End
Directs traffic to AFEs.

4 Application Front End (AFE)
Directs traffic to Application servers.

5 Requests User Data
User's Drive data request goes from the Application to storage.

6 Return Decrypted Data
Send user data to Application.

7 Return User Data
Return user data to user.

8 Return User Data in Encrypted Tunnel
Return user data to user.

DATA DECRYPTION

A Retrieve Data
Gets Encrypted Chunk and Wrapped Key.

B Request Key Unwrap
Wrapped key is sent to KMS.

C ACL Check
Is the requester (e.g. Storage System) authorized to have key unwrapped?

D Send Unwrapped Key
KMS unwraps the encryption key data, which Storage System will use to decrypt chunk.

E Decrypt Data
Storage System decrypts chunk.

Auditing and Access Control for keys data

We complement encryption with rigorous procedures for assigning and removing access to the keys, and logging employee access to the keys and data. We regularly review these procedures and logs to ensure that they are operating in a secure manner and that only the people and applications requiring access are granted it. This process is also audited every year by an independent third party.

We complement encryption with rigorous procedures for assigning and removing access to the keys, and logging employee access to the keys and data.

Google authorizes only trusted individuals to have legitimate access to systems and data repositories containing customer data, including the KMS. This strict authorization extends to job duties including debugging and maintenance activities that might expose decrypted customer data to a trusted employee. Access to these systems is under the umbrella of strict policies that are clearly displayed for employees to read and also in the tools they use. Access to customer data is only allowed for a legitimate business purpose.

To help ensure that only this limited set of trusted employees uses their given access as approved by Google, we use a combination of automated tools and manual reviews to examine employee access to customer data and detect any suspicious events. We strictly enforce our policies for customer data access. We have established an incident response team to investigate violations of misappropriation of customer data. We have established a disciplinary process for noncompliance with internal processes which could include immediate termination from Google, lawsuits and criminal prosecution.

Data on backup media

Google also encrypts all data stored on backup media. Backup media, as noted, are used as a recovery mechanism if there is a failure of the disk data and data needs to be restored. This means that backup media are accessed much less frequently than disks. Each medium contains one or more files, and each file is encrypted with its own unique AES 128-bit file key; these keys are derived by the KMS from a per-file "seed." At backup time, a random seed is created for each file, and the KMS is asked to derive the per-file key by mathematically combining the seed with a "derivation key" known only to the KMS. The resulting per-file key is unique, and is not stored — when it is needed for a restore, it is re-derived from the per-file seed.

The derivation key needed to derive the per-file keys from seeds is known only to the KMS and never leaves it. In addition, only the backup service has permission to ask the KMS to derive per-file keys from seeds. The per-file seeds are stored in the backup system's database. This provides a double layer of access control: (1) only authorized personnel and services may read seeds from the backup system's database, and (2) a further authorization check is required to use such a seed to ask the KMS to derive a per-file key. Both authorization checks must complete successfully to decrypt backed-up data.

In addition, the backup media contain no identifiable information about what is on that medium: all such information is encrypted. An individual who steals a medium with the intent of determining what data is stored on it will be unable to do so.

Finally, the backup system can also back up encrypted files for which it cannot read the plaintext. For such files, it backs up the ciphertext (which it encrypts again via the mechanisms described above) and the wrapped key. At restore time, both are restored, again without the backup system ever seeing the plaintext.

Encryption of Data in Transit

As we've shown, G Suite encrypts customer data stored at rest on both disks and backup media. But we also want to protect your information while it's en route from one machine to another data center, ensuring these data transmissions would still be protected should they be intercepted. Data in transit may be traveling over the Internet between the customer and Google or moving within Google as it shifts from one data center to another.

Data traveling over the Internet

When you use a Google service, your information travels over the Internet between your browser, Google's servers, and, sometimes, non-Google users you are communicating with. In these scenarios, encryption helps prevent hackers from exploiting vulnerabilities in Internet connections to steal your username and password, eavesdrop on your emails, or collect other sensitive data.

Between you and Google

To protect your information, the first step is having a secure browser that supports the latest encryption and security updates. When you're a G Suite customer, we automatically encrypt traffic between your browser and our data centers — whether you're using public WiFi, logging in at the office, or working from home on your computer, phone or tablet.[2] Google websites and properties use robust public key technologies: 2048-bit RSA or P-256 ECDSA SSL certificates issued by a trusted authority (currently the Google Internet Authority G2).

How this encryption works depends on each customer's client configuration. Google servers support ephemeral elliptic curve Diffie-Hellman cryptographic key exchange signed with RSA and ECDSA ("ECDHE_RSA" and "ECDHE_ECDSA"; see Table 2 for details). These so-called forward secrecy methods help protect traffic between customers and Google servers from being intercepted and decrypted by a man-in-the-middle (MitM) attack. In 2011, we announced forward secrecy by default.

Forward secrecy technology helps ensure that information encrypted today is less vulnerable to new methods of breaking encryption in the future.

[2]Sites from Custom Sites require the Admin to install a certificate to enable https.

Forward secrecy technology helps ensure that information encrypted today is less vulnerable to new methods of breaking encryption in the future. With forward secrecy, keys are rotated at least every other day. This limits the impact of a compromised encryption key to information a customer exchanged over a two-day period (instead of what could be several months of data). Without forward secrecy, in contrast, an adversary could record encrypted traffic and store it with the hope of compromising the HTTPS private key at a later date. If they succeeded, they would then be able to decrypt the data.

With forward secrecy, Google servers generate a new Diffie-Hellman public key for each session, sign the public key, and use Diffie-Hellman to generate mutual private keys with the customer's browser. This helps prevent eavesdropping because every session between a customer and Google is encrypted with different public keys. An attacker would have to do two things: capture encrypted traffic ***and*** compromise the temporary private key before it's destroyed.

Forward secrecy also prevents a connection's private keys from being kept in persistent storage. Combined with key rotation, this feature stops adversaries who successfully compromise a single key from retrospectively decrypting data more than two days old.

Between you and non-Google users

We've now reviewed how traffic between a G Suite customer and Google servers is encrypted, but what happens when that customer has business beyond Google?

Google has led the industry in using Transport Layer Security (TLS) for email routing, which allows Google and non-Google servers to communicate in an encrypted manner. When you send email from Google to a non-Google server that participates in TLS, you are protected. We believe increased adoption of TLS is so important for the industry that we report TLS progress in our Transparency Report. G Suite customers also have the extra ability to only permit email to be transmitted with specific domains and email addresses if those domains and addresses are covered by TLS. This can be managed through the TLS compliance setting.

Data moving between data centers

One key advantage for G Suite customers is Google's vast and robust network of Google data centers that spans the globe. Our network is designed to minimize latency and maximize availability, helping to ensure uninterrupted access to your information with no scheduled downtime. In order to achieve this level of performance and conduct upgrades or maintenance, we often move data from one data center to another.

These shifts of data centers are imperceptible to our customers and carried out in a secure manner. Namely, data is always encrypted when it moves between data centers. Connections between internal Google servers are cryptographically authenticated between machines. Certain connections (including those to and from the KMS) are encrypted with a TLS-like proprietary transport protocol that uses AES 128-bit or higher.

Table 2

Encryption protocols and ciphers supported by Google[3]

Protocols	Cipher suites	Signing keys	Hash functions
TLS 1.2	ECDHE_RSA with AES	RSA 2048	SHA384
TLS 1.1	ECDHE_RSA with 3DES	ECDSA P-256	SHA256
TLS 1.0	ECDHE_ECDSA		SHA1
SSL 3.0[4]	RSA with AES		MD5
QUIC	RSA with 3DES		
	RSA with RC4		

[3]This list of protocols and ciphers is subject to change at any time.

[4]Google is working to deprecate old protocols and primitives (such as SSL 3.0, RC4, MD5 and SHA-1) as quickly as users allow. For example, SSLv3 is disabled by default in Chrome 40 and higher. Google Chrome and servers support TLS_FALLBACK_SCSV to prevent attackers from inducing browsers to use lesser protocol versions. Further information is available on the Google Online Security Blog and Google Chromium Group.

Encryption Is Only Part of Our Comprehensive Security Strategy

G Suite customers' data is encrypted when it's on a disk, stored on backup media, moving over the Internet or traveling between data centers. Providing cryptographic solutions that address customers' data security concerns is our commitment. But it's important to note that, while encryption is important and necessary, it's not enough, by itself, to protect your information. Instead, it has to be part of an in-depth, well-organized, and executable security and privacy strategy — like the one we have at Google, which is outlined in the G Suite Security Whitepaper. This comprehensive data protection approach is rare and not typically present in many centralized local computing centers.

Indeed, security has always been central to our daily operations and culture. Our custom hardware and unique data storage architecture are designed with security in mind. We constantly invest in security innovation, employing many highly trained security experts and supporting their extensive and intensive research efforts. We also operate in a manner that helps us quickly respond to newly identified threats and develop better ways to align the protection of customer information with the ever-evolving risk that typifies modern computing. By design our systems restrict access to customer data to only a limited number of individuals and specific applications that require access. For more information on our security practices, please see our G Suite Security Whitepaper.