(IJAER) 2017, Vol. No. 13, Issue No. II, February **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

A REVIEW ON SECURITY OF GOOGLE'S INFRASTRUCTURE & DATA STORAGE

S.B.Rafiah, N.Sreevidya, M.Yellamma

Assistant Professor, Department of IT
Sreenidhi Institute of Science and Technology
Hyderabad, India.

ABSTRACT

Google is the leader of the software community by offering the users many services like Consumer Services and Enterprise level Services.//Google is one of the topmost search engines, is offering the users many services like Consumer Services and Enterprise level Services.// Some of them include Search, online advertisements, Gmail, photos, Cloud computing etc., and also almost every literate without any age barriers makes use of google's services and products. Such gigantic ruler might have followed very strong security techniques. Google's growth itself is an evidence of its secured services. Everyone might want to know how such big organization is providing security to the petabytes of data being generated every second. This paper provides glimpses about its Security design.

Keywords: *Security, Data Storage, Infrastructure, Encryption, Services.*

INTRODUCTION

Most of the People around the world rely on **Google** to search for anything from pin to aero plane. It has become a great success since it's inception on September 15th, 1998. As the Google maintains a huge database of almost everything, Google search is used by nearly 1.17 billion people compared to popular searches like Yahoo, Baidu, Microsoft and Yandex .

Even a school going student or an undergraduate student or a computer illiterate or a software professional or a researcher, whoever may be the person, people of all ages tend to rely on Google for whatever they are looking for. It has very much merged with the lives of people that they have faith in the Google. Google is processing 40,00 search queries every second which equals to 3.5 billion searches per day and 1.2 trillion searches per year. This statistics tells us how much, people rely on Google.

Currently there are 1 billion monthly active users on Gmail account as per the statistics of Feb 1st, 2016. But at some point of time researchers may be worried, at times when they are exchanging their ideas or proposing their research methodologies via Gmail, there may be a doubt that their data may not be secure.

So this paper targets the people who are worried about the security of posting data to Google or Gmail server. The data will undergo multiple stages of checking before it appears to Google user.

This paper is mainly highlighting on how Google is following a tight security check at each and every step of its life starting from a physical level of infrastructure ranging from Server board to operational security where the user is asked to enter username and password. To further increase the security to one step ahead, OTP's are issued to the user.

This paper mainly focuses on how security is provided to Google's Infrastructure in progressive layers from physical security of its data centres to hardware and software that underlie the infrastructure as shown in Fig 1.

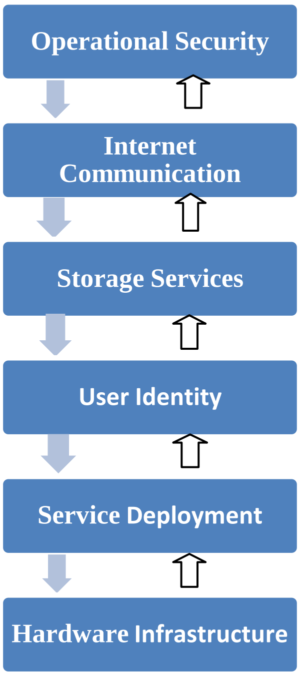

Fig 1: Progressive Layers of Google's Infrastructure

SECURING LOW LEVEL INFRASTRUCTURE

Securing low level infrastructure starts from securing physical premises to securing Boot stack as shown in Fig 2.

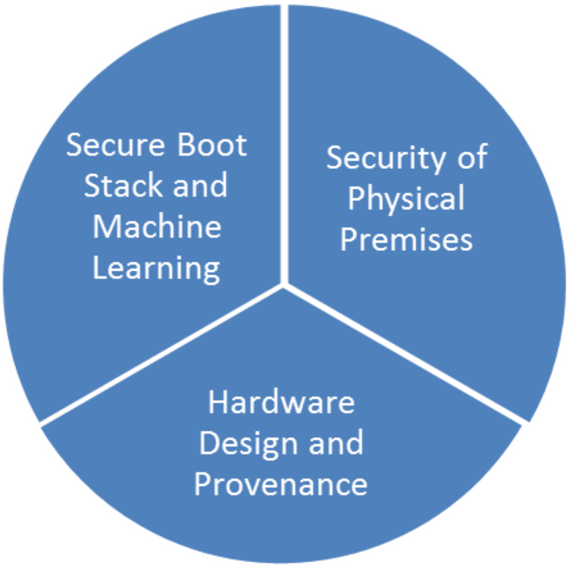
Secure Boot Stack and Machine Learning
Security of Physical Premises
Hardware Design and Provenance

Fig 2: Security at low level Infrastructure

Physical Premises

We shall first talk about security of physical premises. Google develops their data centers[2]with different levels of protection by physically securing them. These data centers are accessed by very few Google employees. Multiple physical security layers are used to protect data center floors. Technologies like Biometric identification, Metal detection, cameras, vehicle barriers and laser based intrusion detection systems are used. In addition to these, it hosts some servers in third party data centers, where google controlled physical security measures are ensured for the top of security layers.

Designing Hardware

Google data centre is comprised of multiple server machines upto thousands which are linked to a local network. The key factor to its security is, it uses its own custom designed Server boards and Networking equipments. A critical examination of component vendors with which they work is done before choosing components. Specially customized hardware security chips are deployed on servers and peripherals which allow secure identification and authentication of legitimate Google devices.

Security measures during Booting

To make sure that the correct software stack is booted, several technologies are used..Low level components like BIOs, boot loader, kernel and base operating system image use cryptographic signatures. These signatures are validated during each boot and update. All the components are constructed and regulated by Google, also hardened by itself.To make sure that the servers run up to date versions in their software stack, detect and diagnose hardware and software problems and to remove any machines from their service, They use automated systems.

SECURITY DURING DEPLOYMENT

To make sure there is security in deploying services on infrastructure and to handle large scales of workload , there will be thousands of machines running the same copies of service for a given service.Cluster orchestration[3] service called BORG controls the services running on Infrastructure.It does not assume any trust between services running on infrastructure.

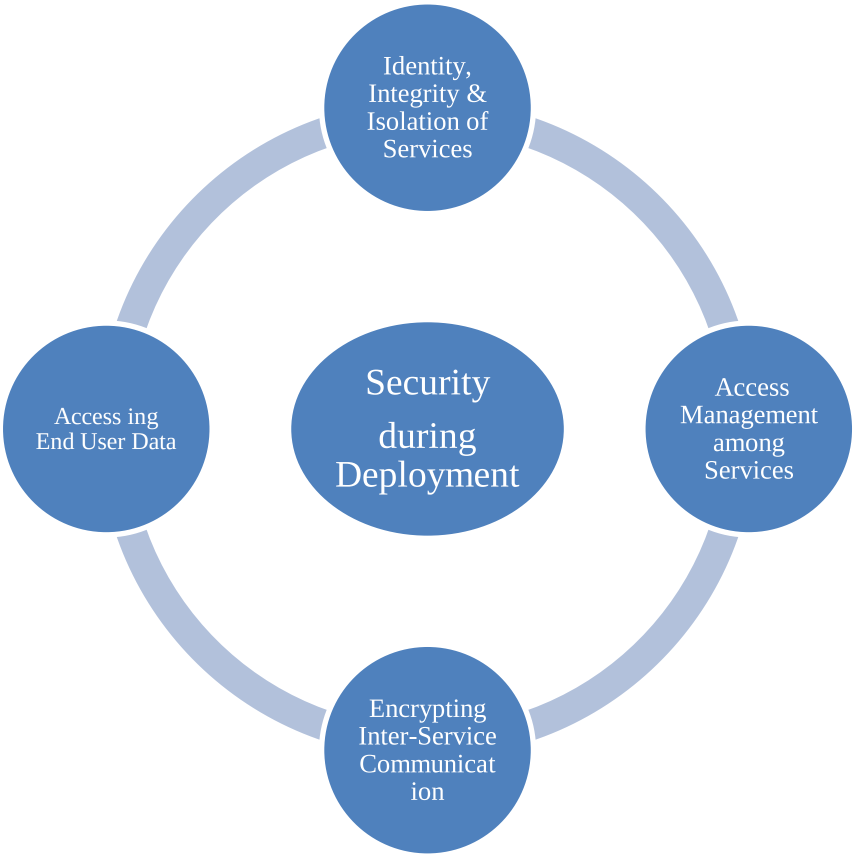
Identity, Integrity & Isolation of Services
Access Management among Services
Encrypting Inter-Service Communicat ion
Access ing End User Data
Security during Deployment

Fig 3: Diagram showing stages of Service Deployment

Fig 3 represents the stages involved in Security during Deployment of Services.

Identity, Integrity and Isolation of Services

Cryptographic authentication and authorization is used at application layer for inter service communication. Internal Network Segmentation or Firewalling are not relied as primary security mechanisms. INGRES and EGRES filtering are used at several points to stop IP spoofing which can aid in increasing the performance of network and network availability.

Each running service is identified by each service account identity. For a service to prove its identity while making and receiving RPC's to other services, it must provide cryptographic credentials.Clients will use these identities to ensure that they are talking to the correct server intended. The source code is stored in a central repository where both current and past versions of service are auditable. Any modifications and code reviews needs inspection and approval . Inspections and approvals from atleast one engineer are other than the owner are needed for any modifications and code reviews.

The system also enforces that code modifications must be approved by owner of the system. .These requirements control the ability of the insider or adversary to perform malicious modifications to source code.A variety of isolation and sandboxing techniques for protecting the service from other services running on same machine are used.

Access Management among Services

The owner of a service can use access management features provided by the infrastructure to specify exactly which other services can communicate with it. The engineers to access services are also issued individual identities, so services can be configured to allow or deny their accesses. The infrastructure is maintaining different types of identities including machine, service and employee in global namespace. The infrastructure provides rich identity management workflow system for these internal identities including approval chains, logging and notification.

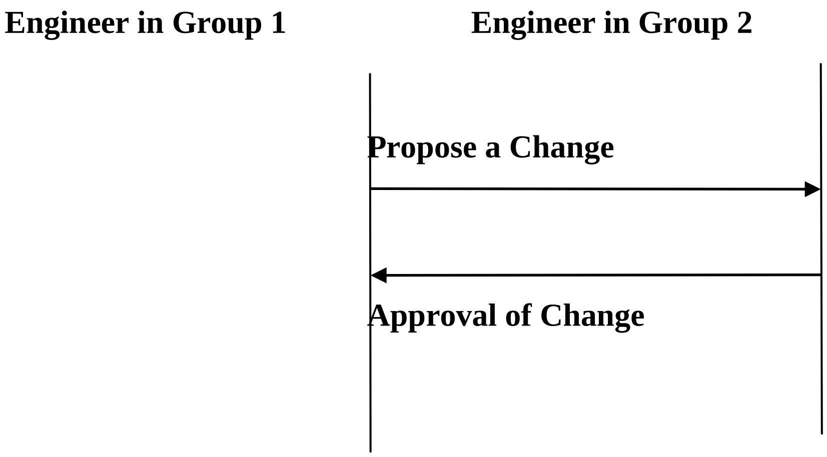

Fig 4: Two party-control system

A Two party-control system where one engineer can propose a change to another engineer(who is also an administrator of the group) which must be approved as shown in Fig 4.. To have secure accessing of management processes , services ranging upto thousands running on the infrastructure support this.

Infrastructure provides some services that can read from central ACL as well as group databases by implementing tailored, fragile access control besides automatic API level access control mechanism.

Encrypting Inter-Service Communication

Infrastructure is providing cryptographic privacy and integrity for RPC data on the network along with RPC authentication and authorization capabilities. For the application layer protocols such as HTTP to gain these benefits, these are encapsulated inside infrastructure RPC mechanisms. The inter-service communication that is encrypted[4]assures to be secure even if there is a tapped network or a networking device is conceded.

Infrastructure is automatically encrypting RPC traffic that is going over WAN between data centres to guard from sophisticated adversaries who are trying to tap private WAN links. Hardware cryptographicaccelerators are deployed to provide default encryption to entire infrastructure RPC traffic inside data centres.

Accessing End User Data

Google's services are programmed to give better service to an end user. For example, an end user having a Gmail account wants to access *address book* needs to call API provided by *contacts service*. Contacts service is designed in such a way that RPC requests from Gmail service or any other particular service that the contacts service wants to allow. With a predefined set of permissions Gmail service can obtain contacts of any user at any point of time

As part of RPC communication, Gmail service is allowed to give “end user permission ticket” in place of a particular end user proving that the Gmail service is servicing a ticket instead of particular end user. This will help the Contact service to send the data to the end user present in the ticket.

A central user identity service is provided whose task is to issue “end user permission tickets”. The service verifies user login and then issues user credentials. Every other request there after should present that user credentials. After the service receives the end usercredentials, it passes these for verification.If the verification is successful,the central identity service gives a ticket for short time period which can be used for RPC’s.

SECURITY MEASURES IN DATA STORAGE

Encryption

To provide security, different types of storage devices like Big table[5], Spanner[6] and a central key management service are used.Many applications access physical storage via these storage devices indirectly. To encrypt data before storing to physical devices, central key management services provides the keys. Automatic key rotation, provision of extensive audit logs are supported. The tickets are used to link keys to particular end users.

As the data is encrypted at application layer , the threats at lower levels of storage such as malicious disk firmware are isolated. Hardware encryption is supported in hard drives and SSD’s to thoroughly track each drive through its life cycle. Cleansing of storage devices consists of two independent verification process. All the storage devices are wiped before they leave their custody. The devices that donot undergo wiping procedure are physically destroyed.

Data Deletion

Deletion begins only after marking specific data as "scheduled for deletion" rather than directly removing the data. This feature allows recovery from unintentional deletion. In accordance with service specific policies, the scheduled data is then deleted. Whenever any user deletes entire account, a notification is sent to services that is handling end user data. Before deleting the data, schedule for deletion is done.

SECURING INTERNET COMMUNICATION

This section discusses how Google provides security among internet and the services. Inter-Service communication's security is not dependent on the network's security as the infrastructure contains huge set of machines interconnected over LAN and WAN.

So the internet's infrastructure is detached to a private address space where protections like defences against DoS attacks can be implemented by allowing partial set of machines disclosed to internet traffic.

Google Front End Service (GFE)

For a service to be present on the internet, it must register with GFE. It makes sure that all TLS connections are aborted with correct certificates by following perfect secrecy. GFE also applies protections against DoS attacks and forwards the service requests using RPC security protocol.

Any internal service willing to broadcast externally uses GFE as a reverse-proxy front end which gives public ip hosting , the things like public DNS name, DoS protection and TLS termination. Like any other service, GFE's run on the infrastructure which can scale to match incoming request volumes.

Protection against DoS

The unmitigated infrastructure of Google helps to imbibe many DoS attacks. Google has a multi-tier, multi-layer DoS protections[7] that reduce the risk of further DoS impact on a service running behind GFE.

After the backbone provides an external connection to one of the data centers, it passes through multiple layers of hardware and software load-balancing. These load balancers are responsible for sending information about incoming traffic to a central DoS service which is running on the infrastructure. Whenever the central DoS service identifies the DoS attack, it will arrange load balancers[8] to drop the traffic i.e, related to the attack. GFE instances report the information about the requests that they are receiving to central DoS service including the application layer information that the load balancers don't have.

Authenticating the User

The central identity service provides Google login page to the end users. Besides asking for a simple username and password, based on risk factors, users will be asked whether they have logged in from the same device or similar location in the past. An authenticated user will be given credentials like cookies and OAuth tokens that can be used at some other time.

While signing in, users can also add additional security mechanisms which include phishing-resistant Security Keys or OTPs. For the benefits to go beyond Google, they have started working in FIDO alliance with multiple device vendors to develop Universal 2nd Factor(U2F)[9] open standard. The devices which are present in the market started using U2F support and extensive web services started using these.

OPERATIONAL SECURITY

Operational Security deals with operating the infrastructure securely. The infrastructure is created securely. The employee's machines and credentials are protected. The threats to infrastructure are defended from both internal and external factors.

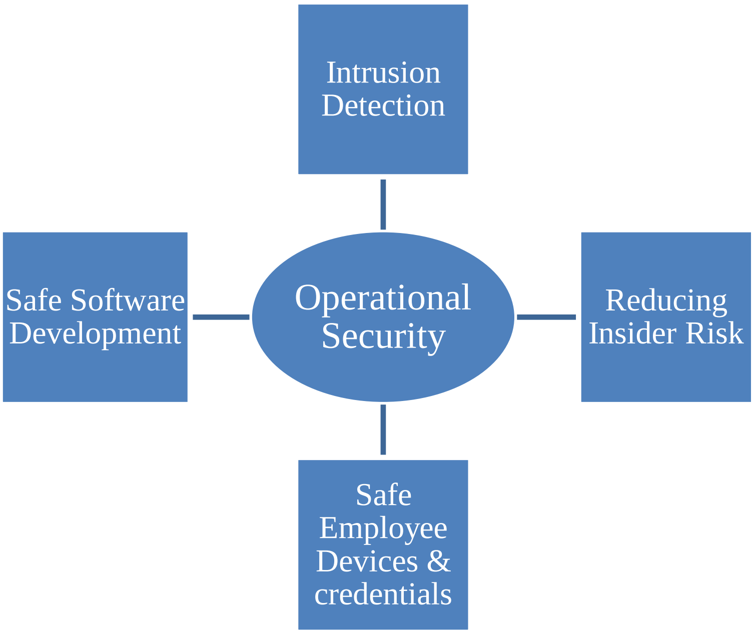
Intrusion Detection
Safe Software Development
Operational Security
Reducing Insider Risk
Safe Employee Devices & credentials

Fig 5: Figure showing the stages of Operational SecuritySafe Software Development

Along with maintaining central source control and 2-party review features, Google also offers libraries which blocks the developers from inducing certain security bugs.

Automated tools detect many security bugs automatically. Despite providing all these, security reviews are done manually. Professional teams which include members from Cryptography, web security and OS security conduct these reviews.

Google runs a Vulnerability Rewards Program[10] where the users are allowed to find out bugs in their applications or infrastructure, where they are paid. Google does lot of effort in finding 0-day exploits and other security issues in the operating systems they use.

Employee Devices and Credentials at safe

Google to maintain the infrastructure secure takes measures in protecting employee’s devices and credentials and monitors the activities for any vulnerabilities.

- To guard against sophisticated phishing, phishable OTP factors have been replaced with U2F security keys for their employee accounts.
- The client devices that employees use are also monitored.
- The client devices are always updated with security patches.Some dedicated systems are used to scan certain user installed apps, downloads, browsers, extensions and content from web.

  The access privileges are not directly granted by mere presence of the device on the corporate LAN but also application level access management controls are used. These controls expose internal applications to only specific users when they come from a certain device and from expected networks.

Reducing Insider Risk

The activities of employees who have been granted administrative access are monitored actively. Privileged access for particular task can be done in a safe and controlled way with automation.The two party approvals are required for some actions. Some APIs allow debugging without providing sensitiveinformation. The security team monitors the access patterns and fins out the unusual events.

Intrusion Detection

Host-based signals on individual devices, network signals and signals from infrastructure services are integrated into well-equipped data processing pipelines by Google.

These pipelines are built with rules and machine intelligence to warn the engineers from any incidents.The incidents are responded by various teams such as investigation and incident response team. To measure and improve the effectiveness of their detection and response mechanisms, Red team exercises are performed. They are monitored and responded 24 hours a day, 365 days a year.

ACKNOWLEDGEMENT

We would like to thank Google for inspiring us in writing this paper by motivating us and we have tried with all our efforts to retrieve the best of Google's white paper on “Google Infrastructure Security Design overview”. So that we can reach any software literate, about how Google is

following procedures to provide the best of security measures to its end users without any compromise.

CONCLUSION

No wonder Google is the top rated search engine in the market as it uses tight security at every stage. This paper discussed on how the security goes from the physical level of infrastructure to Operational Security. They go through rigorous checks performed both manually and through automation to avoid any compromise at any stage. They build their own Server Boards, microchips and their own devices to ensure that data is not compromised at any stage of its processing.Because of the norms it follows in providing security in its design, it has become the giant of all search engines.

REFERENCES

1. Googles white paper on “Google Infrastructure Security Design overview” “https://cloud.google.com/security/security-design/

2. Physical security of our data centershttps://goo.gl/WYlKGG

3. Design of our cluster management and orchestrationhttp://research.google.com/pubs/pub43438.html

4. Storage encryption and our customer facing GCP encryption featureshttps://cloud.google.com/security/encryption-at-rest/

5. BigTable storage servicehttp://research.google.com/archive/bigtable.html

6. Spanner storage servicehttp://research.google.com/archive/spanner.html

7. More about DoS protection best practices on GCPhttps://cloud.google.com/files/GCPDDoSprotection-04122016.pdf

8. Architecture of our network load balancinghttp://research.google.com/pubs/pub44824.html

9. Combating phishing with Security Key & the Universal 2nd Factor(U2F) standardhttp://research.google.com/pubs/pub45409.html

10. More about the Google Vulnerability Rewards Programhttps://bughunter.withgoogle.com/