Sonian Archiving & e-Discovery Service

Application & Data Security

sonian

Contents

APPLICATION & DATA SECURITY

1. Objectives

The purpose of this document is to address the following areas:

- Data in transit: endpoints/trust, connections, hops, etc.
- Data at rest: digesting and encryption, encryption key management, etc.
- Application security: authentication and trust, connection security, etc.

This document summarizes the secure data storage management details of the Sonian Archive Service.

1.1 Audience

This document is written for the Customer IT administrator, the Chief Security Officer or any other stakeholder who needs to know how the Sonian Archive Service secures customer data stored in a cloud-computing environment.

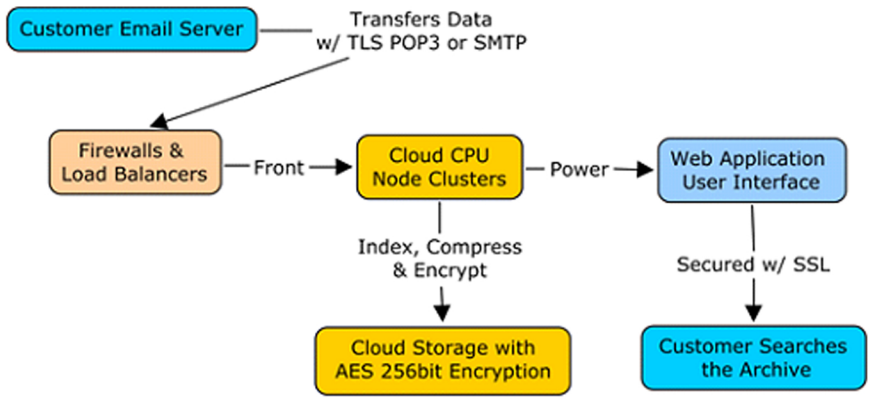
1.2 Architecture Overview

One component of the Sonian Archive Service's core value proposition is to provide data security in two forms:

A) Privacy, which means no customer data is accessible except for authorized customer representatives (aka "admins" or "compliance officers" or "end users")

B) Resiliency, which means data stored in the archive has a very high SLA for always being accessible and no data items should ever "get lost

Figure 1. High-Level Sonian Archive System Overview



The Sonian Archive Service's main components include firewalls, load balancers, cloud compute nodes (CPU) and cloud storage.

Customer email servers communicate with the Sonian Archive Service with TLS encrypted connections. Customers can choose to use either SMTP or POP3 for data import to the archive.

Firewalls protect each compute node in the cluster.

Load balancers distribute inbound connections to available compute nodes for performance and scale.

Cloud CPUs are clusters of compute nodes that provide different functional services to the full data processing path for all archived information.

Compute node clusters are used for:

- Enterprise messaging service to coordinate workflows between components
- Data import service to receive data from customer import
- Processes that index, compress and encrypt data
- Database sub-systems to store original data (in an immutable format), meta-data (light-weight descriptive data), index data (full text search) and relational data (SQL)
- Application servers that allow the customer to search the information in the archive

2. Sonian Archive Service Application & Data Security

The Sonian Archive Service uses cloud storage repositories to securely store data. The chain of custody from the customer site to Sonian Archive Service cloud storage is preserved by TLS encryption for data in transit and 256-bit AES encryption for data at rest.

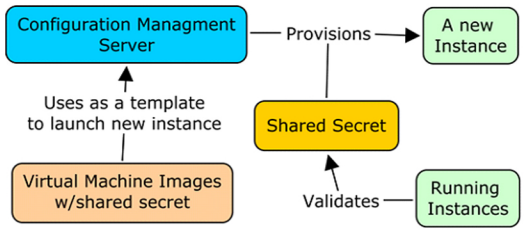
The Sonian Archive Service provides SMTP & POP3 over TLS. All compute instance nodes are automatically configured from a centralized system that manages and authenticates each processing node. All disks and data are encrypted within the cloud infrastructure environment so that there is no data visibility to cloud hardware technicians. All search databases are in separate locations by customer so there is no data co-mingling. Likewise, all original data objects for each account is stored in a per-customer location on cloud storage, so there could not be an accidental retrieval of incorrect email data from one customer to another.

Sonian Archive Service systems administration personnel use Multi-Factor Authentication (MFA) devices to gain access to cloud compute configuration systems. MFA is security layer that offers enhanced control over the cloud infrastructure. MFA requires a valid six-digit, single-use code from a physical authentication device to access management consoles, security access identifiers, compute settings and other critical functions.

2.1 Operational Security

2.1.1 Configuration Management

Figure 2. Cloud-Compute Configuration



The Sonian Archive Service cloud compute virtual machine images are packaged with a configuration management client and a pre-shared configuration management secret. The virtual machine image is packaged, encrypted & signed and placed on cloud storage. The images are private and only bootable by the Sonian Archive Service. This preserves the integrity and control of all processes (the programs that touch customer data) that launch automatically.

The cloud-compute nodes are automatically configured by sharing a secret between the Configuration Management Service and the new nodes being booted. Communication between the new node and the configuration management server is over TLS. If the secret does not match, the new node cannot be configured and no information is shared with the new node. After the node has been validated, the configuration management service instructs the node which software, services and configuration changes are required to assume its given role (backend, web application, SMTP, POP3, etc.) in the cluster. Once the node has been configured, the Configuration Management Service continuously enforces the original configuration and software settings to ensure no manual changes have occurred.

Multi-factor authentication service and devices are used for system administration login and access to specific settings and sensitive (operational) data information screens.

2.1.2 LUKS Disk Encryption (Swap & Data Partitions)

The Sonian Archive Service encrypts cloud compute node virtual disk drives with LUKS (standard Linux disk encryption). On boot, the configuration management system places a LUKS encryption key(s) on disk for the period needed to mount the encrypted volumes and swap.

The purpose of encrypting the disk drives and swap on every node is to protect against inadvertent data leakage to other cloud compute nodes. Encrypting disk and swap space with LUKS ensures the Sonian Archive Service compute volumes and nodes from being seen on other nodes and foils forensics tools.

In addition, the underlying cloud infrastructure (at the virtualization layer) automatically "zeros out" each block on cloud storage volumes prior to use. This is an additional layer of protection.

2.1.3 Encrypted Inter-Node Communication

The communication between all nodes is protected by TLS encryption. Each node has an internal, private IP address. In addition, some nodes that need to communicate with entities outside the cloud get a public IP address.

2.1.4 SSH Keys

The Sonian Archive Service only allows SSH keys as the method to gain access to a cloud compute node. Passwords are not used for logins.

2.2.1 Securing Data Imported Over The Wire

The Sonian Archive Service recommends all customers use encrypted SMTP/TLS for data import into the archive. For POP3, we recommend using POP3/TLS connections.

In general, the customer is responsible for initiating the SSL connections.

- SMTP/TLS: The sending server (the customer's email server) must start the SMTP conversation and force secure transport. The Sonian Archive Service will accept a secure SMTP connection and the SMTP conversation will be encrypted.
    - Each email server technology has customer IT administrator control panels to enable SMTP secure transport. This is typically called "SMTP/TLS" or "Enable Start TLS" or "SMTP over TLS." The Sonian Archive Service will respond appropriately to any of these secure SMTP methods.
- POP3/TLS: The email server hosting the POP3 mailbox for data collection can typically support an inbound TLS connection. The Sonian Archive Service can be configured to use TLS so that no data is sent over clear text.
    - The customer will need to enable the TLS option for the POP3 protocol on their email server.
    - The Sonian Archive Service "Data Collector" configuration screen has an option to force TLS connections and the customer can test the TLS connection before importing data.

2.2.2 Securing Bulk Data Import

The Sonian Archive Service supports a data import method called "Bulk Data Importing," which allows physical media to be shipped to a processing facility and the files on the media (typically USB or eSata drives) will be imported into the customer's archive account.

The customer can choose to encrypt the files prior to shipment and then notify the Sonian Archive Service support personnel of the decryption password so that the files can be accessed.

Sonian also supports the transfer of bulk data imports utilizing secure cloud object storage services, including AWS S3, Azure Blob, Rackspace Cloud Files and Softlayer Object Store.

2.2.3 Securing Data Exported From The Archive

From time to time, customers may want to export their data from the archive. This is a customer self-service function initiated from the web application.

The customer can select items to export and choose a format of PST, File, PDF or HTML. The customer will receive an email alert when the export file is ready to download. The export file will also have a MD5 hash generated to be used for verification and authenticity that the file hosted on the server is the same file downloaded to a desktop computer.

2.2.4 Data Disposition After Canceling Service

If a customer desires to cancel the service, their data can be electronically shredded or returned on physical media.

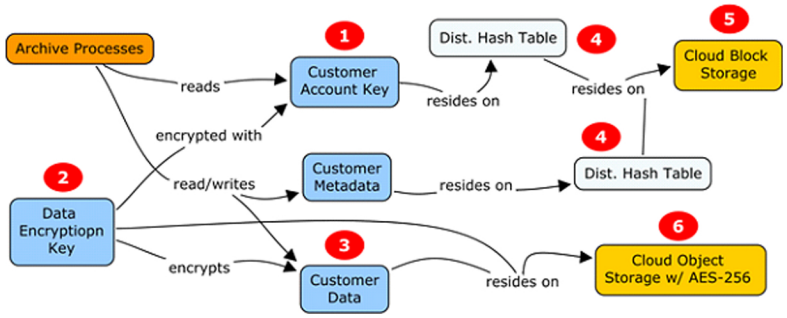
Electronically shredding removes encryption keys and all customer data is deleted.

Returning data on physical media is a process similar to physical media import, except in reverse. The customer can request a shipment from the Sonian Archive Service of a USB or eSata drive(s) with the archive files in an industry standard format. There is an additional fee for the bulk data export service.

2.3 Storage Security

2.3.1 Encryption

Figure 3. Data Encryption



When a new customer account is created, a random keystore password (1 and 2) is created using the Java Cryptographic Extension (JCE) APIs and a third-party security provider library.

The hash of this password is kept in a database (4). This password hash is used to encrypt access to the customer's data key in the keystore. The keystore key is in a Dynamo-like meta-database (5). It is a random keystore key and not a hash of a password. The keystore is kept on S3 (6). The customer-specific keystore password can be changed at any time using this method. Data objects do not need to be re-encrypted when the password hash changes.

Sonian generates a per-account private data encryption key when the customer's account is created on our system. This private key is used for encrypting and decrypting objects on S3 (6). The private data key is only accessible using the customer's password hash from the database and the customer's keystore from AWS S3.

- On account creation, a unique, random Account Key is generated, conforming to NIST 800-90 compliance (1).
- Each account has its own Java Keystore on the cloud storage service. The Java Keystore is similar in concept t to an OS X KeyChain. (1 and 2).
- Each Java Keystore contains one AES-256 key locked by the Account Key (1 and 2).
- The Account Key is used by the server processes that manage data to access the Encryption Key via the Java Keystore.
- The Encryption Key is used to do the actual encryption/decryption of an account's data.
- Should the Account Key change, the Encryption Key is re-stored in a new Java Keystore on cloud files (The data doesn't need to be re-encrypted, as only the password to the Encryption Key has changed.)

This design supports the need for secure practical data encryption to support the needs of programmatic access to the Encryption Key to import, export and have a UI with access to the data.

2.3.2 Digests

A digest of the appropriate parts of the document's meta-data and body is maintained. This digest is stored in the meta-data and also in the audit trail.

2.3.3 Strategy For Key Store Breaches

A security breach is highly unlikely, but if such an event were to occur, the customer data can be "re-keyed" quickly and immediately.

2.3.4 Data Resiliency

Data resiliency is based on cloud storage resiliency SLA of 99.999999% durability guarantee. In addition, all customer meta-data and indexes can be re-built from the source data stored on cloud files.

The Full-Text Search Clusters and Meta-Data Database Clusters are based on "Dynamo," which is described as ".... [a] technology designed with the ability to trade-off cost, consistency, durability and performance, while maintaining high-availability." [Source: Dynamo Whitepaper http://goo.gl/zOG7]

Systems architected with dynamo principles are highly durable and take advantage of cloud computing's "scale-up/scale-down" utility model.

Where required for uptime SLA and performance, the Sonian Archive Service technology uses cluster concepts to achieve both goals.

2.4 Audit Trail

Figure 4. Audit Trail Data Flow

Each user action and system function in the application is stored in the audit trail. For customer benefit, some audited events are viewable in a password-protected RSS feed.

2.5 Conclusion

The Sonian Archive Service is a state-of-the-art system designed to manage email content in a cloud-computing environment. The benefits of cloud computing include: no single point of failure, an extremely high data storage reliability SLA and a data-security model that ensures privacy and resiliency.