Use Case

AGENCY PROTECTS AGAINST INFECTED FILES IN DATA LAKE BUILT ON AMAZON S3

CLOUD STORAGE SECURITY

CloudStorageSec.com

AT A GLANCE

Organization Overview

- Agency processes critical data to aid in the defense of national security and interests both domestically and abroad
- Ingests dozens of TB of data each month from vendors and data brokers for processing within the agency
- Data ranges from millions of small files to large classified files stored in a data lake
- Processed files stored within the agency data lake are used internally and by external partner agencies

Key Takeaways

- A scalable antivirus solution specifically designed for AWS was needed
- File sizes range from a few KB to over 200 GB in size
- Complex and proprietary files types used for analysis needed to be supported for scanning and cloud detonation
- Integrated as part of a unified security operations center using Proactive Notifications and AWS Security Hub

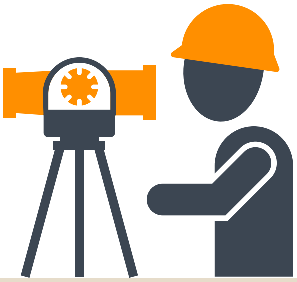
THE CHALLENGE

Like many public sector agencies, this organization uses Amazon S3 to ingest and store large amounts of data from third party vendors and data brokers for processing.

During a security audit, it was determined that the ingestion of TB of data each month from outside sources is a major attack vector. There was no assurance that the data has been scanned for malware and viruses before it was processed and accessed by agency staff or shared externally with partner agencies.

Recognizing that virus scanning for objects in Amazon S3 isn't provided by AWS-native services, the agency needed a solution to scan the data for potential threats.

The initial use case was to scan 5-10 TB of surveyor data on a monthly basis, eventually leading to a range of proprietary file types over 200 GB in size.



REQUIREMENTS FOR SOLUTION

A solution that would meet the organization's needs today and grow with them far into the future was needed. Originally the team researched a homegrown solution using ClamAV® and AWS Lambda but quickly realized the limitations of that solution would not suffice. Their requirements included:

- Deployment within their AWS account, guaranteeing no one outside of the organization could gain access to sensitive data
- Large file scanning with the ability to scan large volumes of files at once
- A tool that catches infections at the data ingestion layer without disrupting their workflow
- Scalability allowing for management across hundreds of AWS accounts, thousands of Amazon S3 buckets and users from across different departments
- Easy-to-use, comprehensive interface eliminating the need to build and maintain reports or use multiple security tools
- The ability to automate scanning on a quarterly scheduled basis to meet federal security compliance requirements
- Bucket assessment to see which buckets are public, unencrypted, or misconfigured
- Integrations for AWS Security Hub

10 TB	5 TB	~~Misconfigurations~~	Integrations	1,000s of Dollars
Avg. Amount of Data Scanned Monthly	Largest Scannable File Size	Security Assurance thru Bucket Attribute Reporting	Critical Findings Published to AWS, Security Hub, SIEM	Est. Cost Savings from using Smart Scan

WHY ANTIVIRUS FOR AMAZON S3

Antivirus for Amazon S3 by Cloud Storage Security is a modern, serverless solution that enables organizations to scan high volumes of data quickly without delaying data ingestion and data processing workflows.

It is deployed in-tenant so the data never has to leave the agency's highly secure AWS environment.

The solution offers the premium Sophos Antivirus Dynamic Interface scanning engine, allowing the agency to scan a wide range of file types (including a large variety of proprietary file types) and can also scan very large files up to the 5 TB maximum file size permitted by Amazon S3.

By leveraging Sophos for scanning, the agency also receive the latest updates from global threats and malware.

Scalability for both users/groups as well as agents for scanning large volumes of files is easily configurable.

Additionally, the solution integrates as part of an organization's security operations center using AWS Security Hub, Proactive Notifications, and SIEM integrations.

THE SOLUTION IMPLEMENTED

Cloud Storage Security supported the agency during a 30-day POC delivered in AWS Marketplace. During the POC, we

- helped the agency deploy Antivirus for Amazon S3 to meet their security requirements
- validated that Antivirus for Amazon S3 can support scanning their required file types and sizes
- assisted with scalability testing to ensure that scanning for large data sets was performed within expected time frames.

The agency was able to move from POC to production scanning in 30 days, tripling scan volume in the second month.

Additionally, the agency configured Proactive Notifications in order to be alerted anytime an infected file was found. They integrated the alerts into their SIEM system creating a centralized view as part of their unified security operations center.

To create and maintain a clean, secure data lake, the agency implemented two Antivirus for Amazon S3 scanning models - event-based and retro. To optimize cost, they enabled Smart Scan functionality, which allows for agent scaling. To address problem files, they implemented alerts and can employ static and dynamic analysis.

Event-based Scanning

The main scanning model used by the agency is event-based scanning. This is due to the ongoing ingestion of files, which generally arrive in bursts between 500 GB and 1 TB.

The agency utilizes our 2 bucket system document flow to quarantine infected files in a separate bucket and guarantee that no one has access to the file until it has been confirmed to be safe.

Retro Scanning

To meet federal security requirements, a quarterly scheduled scan is performed on all existing data stored within the data lake.

This is to ensure that all data is clean and that files are tested against any new virus definitions that have been released since the last time they were scanned.

Agent Scaling

Because data is ingested in bursts, the agency enabled our Smart Scan functionality, which scales agents down when scanning activity isn’t at a high volume, and scales agents up when high scanning volumes are detected.

In the long run, the use of Smart Scan saves thousands of dollars because scanning agents are only activated when needed.

Problem Files

When an infected file is detected, the agency is alerted within seconds via Proactive Notifications and a SIEM integration, which allows them to respond quickly.

Using Static Analysis, the team is able to analyze the file and generate information about the threat that has been found. If further investigation is required, Dynamic Analysis via cloud detonation is available to explore and verify the threat inside the problem file.

THE RESULTS

The team evaluated products in AWS Marketplace. The agency selected Antivirus for Amazon S3 because of its ability to scan large multi-GB files efficiently utilizing the Sophos virus detection engine and the ability to customize the deployment to meet their security requirements. The product's scalability and support for scanning files up to several TB in size also means that it can support future projects that will ingest even larger files and data sets.

Antivirus for Amazon S3's ability to automate the scaling of scanning agents through Smart Scan ensured that TB of data can be ingested in bursts and scanned for potential threats without delaying data processing.

- An average of 10 TB per month scanned, with the largest month so far being over 50 TB
- Federal security compliance requirements met through quarterly scheduled scans
- Improved S3 bucket configuration and security across thousands of buckets by having a unified view of all public/private, encrypted and misconfigured buckets
- Unified view of threats through use of Proactive Notifications and SIEM integration