Scanning Amazon S3 Bucket Using ClamAV: Serverless Solution

**PREPARED BY**

Vivian A Lem

# Introduction

Amazon Simple Storage Service (Amazon S3) is a highly scalable object storage service that allows organizations to store and process data. Because of its flexibility and ease of use, it has become the most used service in many applications hosted on Amazon Web Services (AWS).

Wide variety of solutions use S3 as a storage option. The data coming from third parties or uploaded from the internet may be infected with malware, viruses, ransomware, trojan horses, and more.

To safeguard Organizations and end users using this object/files data coming for data source as S3, we needed a solution which will scan each file upon uploaded to S3 bucket

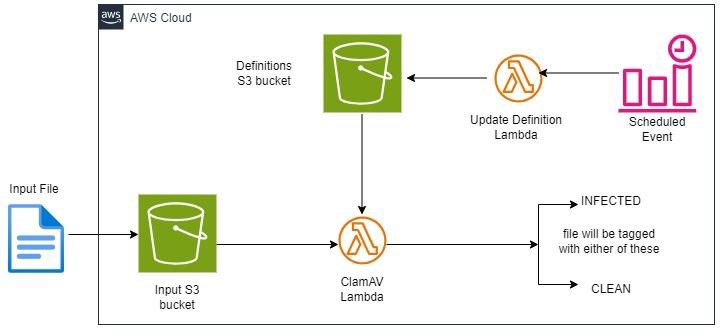
# Purpose

The purpose of this document is to outline the technical design for implementing object scanning upon upload in Amazon S3. This feature enhances security by automatically scanning objects for potential threats or policy violations as they are uploaded to the S3 bucket.

# Scope

This design covers the integration of an object scanning mechanism into the S3 upload process, ensuring that all objects undergo a security check before being stored in the S3 bucket.

# Technical Architecture



* Amazon S3 Bucket

Amazon Simple Storage Service (Amazon S3) is a highly scalable object storage service that allows organizations to store and process data.

In the provided solution we will be having two separate S3 buckets.

1. Input S3 bucket - This bucket will be used to store the object files
2. Definition S3 bucket - This bucket will be used to store virus definition files.
   * Object Scanning Service Lambda

ClamAV lambda will be responsible for scanning objects for threats or policy violations. This lambda service leverages virus definition files from Definitions S3 bucket.

Upon file upload to input S3 bucket, S3 event triggers ClamAV Lambda. This Lambda scans the files against the virus definitions files retrieved from Definition S3 buckets. Upon scan complete, the file will be tagged either CLEAN or INFECTED depending upon scan result.

* + Event Trigger

An event trigger configured with cron job (set to execute every 3 hours) to invoke the Update Definition Lambda

* + Update Definition Lambda

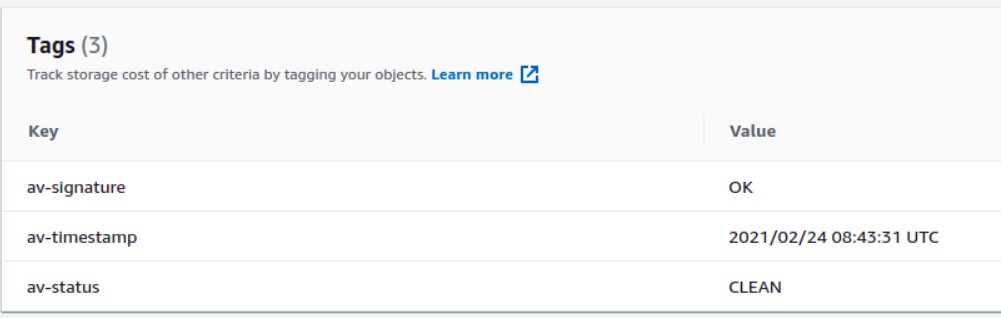
Update definition lambda will be responsible for updating definitions in Definition files in S3 bucket. This lambda gets triggered with a CloudWatch event using cron job executing every 3 hours.

# How it works

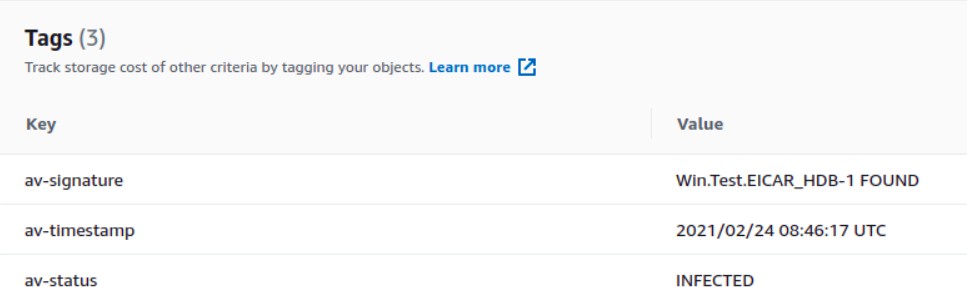
* + Each time a new object is added to a bucket, S3 invokes the Lambda function to scan the object.
  + The function package will download current antivirus definitions from a S3 bucket.
  + The object is scanned for viruses and malware. Archive files are extracted, and the files inside scanned also.
  + The object's tags are updated to reflect the result of the scan, CLEAN or INFECTED, along with the date and time of the scan.
  + If the file is INFECTED, the bucket policy restricts access to the file, and you can't open or download it.

# How looks result after scanning

The result can be like this



* + in the CloudWatch logs



# Prerequisites:

1. Terraform installed on local machine
2. AWS cli need to be installed
3. AWS configuration should done on local

# How to spin up infrastructure

1. Git clone the repo
2. cd clamav-scan
3. Create main.tfvars file with correct aws\_region and buckets\_to\_scan with list of buckets that ClamAV will be check
4. Run terraform init
5. Run terraform plan
6. Run terraform apply