Using AI to Detect Malware in Object Storage

Author: Matthew Battagel, Supervisor: Theodoros Spyridopoulos

Acknowledgments -

I would like to extend my sincere gratitude to my supervisor Theo, family, friends, colleagues, and girlfriend Lois for their unwavering support and encouragement during my research project. The guidance and expertise provided by my supervisor and colleagues were critical in the shaping and success of the project. The support and constructive feedback from all parties helped me to refine my work notably Lois's constant motivation and dedication were a great source of strength throughout the project. I am truly grateful to all of them for their contributions.

Abstract

Lorem Ipsum

Contents

1. Introduction

Overview

The exponential growth of data generation has made data storage an increasingly important aspect for both individuals and organizations alike. Object storage has emerged as a promising solution due to its ability to store vast amounts of unstructured data in a cost-effective and scalable manner. Unlike traditional storage techniques, object storage stores data as objects with related metadata and unique identifiers, allowing for efficient and cheap storage within buckets.

One of the most widely used object storage platforms is Amazon S3, which provides a highly scalable and reliable solution for storing data. However, an open-source alternative called MinIO has emerged as a promising contender, providing similar features to Amazon S3 while giving customers greater control over their data. MinIO is written in Go and is available for free under the Apache License 3.0 or, for commercial and enterprise purposes, at a reduced cost compared to Amazon S3. (**?**). MinIO offers a wide range of features, including high performance, data replication, encryption and erasure coding (MinIO, 2023). Most importantly, MinIO is designed to scale out horizontally to ensure that it can handle the demands of large-scale applications.

Scalability is made simple by allowing multiple types of hardware platforms to work together in separate nodes each with their own compute and storage. This is extremely attractive for customers who want to utilises their existing hardware without being tied down to a specific provider. This also applies for customers looking to migrate their data from Amazon S3 to cheaper solution without compromising on the high performance, reliability and scalability of the S3 platform.

While MinIO is a great alternative to Amazon S3, it does not offer any form of malware detection integration. This could put customers off from choosing MinIO as a viable platform to migrate to from Amazon S3 or leave existing users data vulnerable to malware attacks. This project aims to address this issue by integrating a malware detection system into MinIO. An important goal for the is to negatively impact the scalability or performance as little as possible so that MinIO is still an effective alternative to Amazon S3.

Motivation

Due to the high amount of unstructured data expected to be both written and read to the object store, there are increased risk of encountering malicious files. Therefore malware detection within object storage is crucial in modern cloud storage scenarios. Most popular off-the-shelf object storage platforms, such as AWS, already have integrated third-party antivirus software, such as ClamAV and Sophos (Srinivasan *et al.*, 2022), to mitigate security risks. MinIO on the other hand is vulnerable to malware attacks as it currently does not have any native antivirus integration. This forces customers who require complete virus protection to either not use MinIO or to use potentially costly third-party software. As antivirus scanning is inherently resource intensive, if the software is integrated incorrectly, it could reduce the ability for the storage solution to scale horizontally which negates one of the major benefits of object storage. The purpose of this project is to implement malware detection within MinIO while being mindful to not impact the scalability or performance of the platform.

Project Aims

This project aims to supply an end-to-end solution for detecting malware within the MinIO object storage platform with three main requirements. The solution should be able to detect the latest uploads to the object store and scan them. It should be able to perform this function without significantly impacting the performance of the object store. The solution should also scale alongside MinIO to ensure that it does not bottleneck the object store at high loads. These main three aims can be quantified so that an accurate evaluation of the solution can be made at the end of the project.

- Detection of malware within the object store should match 100% of the malware detection rate as the stand-alone antivirus.
- Performance of the solution to be within 10% of the performance of the stand-alone object store, MinIO.
- Retain the previous metric while both increasing the available resources and changing the platform.

Important outcomes / targets

2. Background

Amazon S3 Malware Detection

As MinIO’s largest competitor, this project draws a lot of inspiration from Amazon S3s integrated malware detection blog page (Srinivasan *et al.*, 2022). The blog explains Amazons current approach for managing malware detection within their service. Amazon S3 uses a combination of ClamAV and Sophos as their third-party scanning engines due to their out-of-the-box nature. Amazon then gives you the option to use either of these engines or both. The blog goes on to describe the three main interaction mechanisms that Amazon S3 uses to flag files for scanning. Firstly, an API endpoint would be provided to handle all uploads. This forms a queue of uploads which are then scanned before entering the bucket. Next, event-driven scanning is used keep track of all regular file uploads. The antivirus will then scan each file after they have been written to the bucket. Finally, retro-driven scanning is used to scan all existing files within the bucket. The user then has the flexibility to define what types of files should be scanned including defining time windows. This blog has given some useful methodologies of how to keeping track of both incoming and previously scanned files. Creating a system that can match these methods is important for offering a matching level of scalability and security within MinIO.

3. Specification and Design

Specification

Architecture

Design 1

Design 3

Design 2

Design 4

Optimal Design

4. Implementation

Event Bus

Packaging

5. Results and Evaluation

6. Future Work

7. Conclusions

8. Reflection on Learning

9. Appendix

References

MinIO. 2023. Minio object storage, Available at: https://min.io/. [Accessed: 15 Mar 2023].

Srinivasan, G., Eidelman, A. and Casmer, E. 2022. Integrating amazon s3 malware scanning into your application workflow with cloud storage security, Available at: https://aws.amazon.com/blogs/apn/integrating-amazon-s3-malware-scanning-into-your-application-workflow-with-cloud-storage-security/. [Accessed: 21 Mar 2023].