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

Motivation

Project Aims

Important outcomes / targets

2. Background

Malware

Malware is a type of software designed to harm or exploit computer systems, networks, and users. It includes a wide range of harmful programs, such as viruses, worms, Trojans, ransomware, spyware, and adware. Malware can be used to steal sensitive information, gain unauthorized access to systems, damage or destroy data, and perform other malicious activities. Malware can be distributed through various channels, such as email attachments, malicious web-sites, software downloads, and infected removable media. It is a serious threat to computer security and can cause significant damage to storage devices if left untreated.

Malicious files can be particularly risky for storage devices in cloud environments due to their shared nature. In a cloud storage environment, multiple users and applications can access and store data on the same physical storage infrastructure. This means that an infected file can quickly spread and infect other files and users, compromising the security and integrity of the entire system. Moreover, cloud storage providers may not be able to isolate and contain malware as easily as with traditional storage systems. As a result, cloud storage users need to be extra vigilant and take proactive measures, such as implementing antivirus software, regular backups, and secure access controls, to mitigate the risks of malware infections.

Anti-Virus

In the context of modern IT solutions, ensuring robust security measures is crucial to guarantee reliable operations and safeguard data against threats. Companies can face severe penalties, up to XXXX, for failing to comply with data protection regulations in the event of a breach. Therefore, it is imperative that IT products offer comprehensive security features, mainly including built-in antivirus capabilities, to protect against malicious attacks and prevent data loss.

Modern Anti-Viruses mostly work using a combination of signature-based and heuristic detection methods. Signature-based detection works by comparing the file being scanned against a database of known malicious file signatures. If a match is found, the file is flagged as malicious. This method is effective at detecting known malware, but it is not effective at detecting new or unknown malware. Heuristic detection, on the other hand, works by analyzing the behavior of the file being scanned and comparing it against a database of known malicious behaviors. This method is more effective as it looks at practices used by malware, such as common imports, and therefore has potential to detect previously unknown malware.

Artificial Intelligence

Object Storage

Object storage is a type of data storage architecture that manages data as discrete units known as objects. Each object contains data, metadata (information about the object), and a unique identifier that enables it to be located and retrieved. Object storage systems are designed to handle large amounts of unstructured data, such as files, images, videos, and other multimedia content and have become increasingly popular in recent years with a compound annual growth rate of 13.6% (Ot, 2022).

Unlike traditional storage architectures like file or block storage, object storage does not organize data in a hierarchical directory structure or use fixed-sized blocks. Instead, it allows data to be stored and accessed independently of the underlying physical storage infrastructure. This makes it easier to scale storage capacity and performance, as well as to implement features like data replication, versioning, and encryption. Object storage is often used in cloud computing environments, where it is a popular option for storing and managing data in distributed systems.

MinIO

MinIO is a high-performance, open-source, distributed object storage system that is designed to be cloud-native. It provides Amazon S3-compatible API for developers to build cloud-native applications. MinIO is designed to scale horizontally and can be deployed on a wide variety of hardware and software platforms (MinIO, 2023). It is written in Go and is available under the Apache License 2.0 (MinIO GitHub Repo, 2023). MinIO has several features that make it an attractive choice for cloud-native applications. It is fault-tolerant, with data being automatically distributed across multiple drives and servers (erasure) to ensure high availability. Additionally, MinIO is highly performant, with a focus on optimizing for SSDs and NVMe drives (MinIO, 2023).

Currently, MinIO does not provide built-in antivirus capabilities. This means that users must rely on third-party antivirus software to protect their data from malware. However, this can be a challenge for users who want to use MinIO in cloud environments, as antivirus software is often incompatible with cloud-native applications. This is because antivirus software is designed to run on a single machine and is not designed to scale horizontally. As a result, antivirus software is not well-suited for cloud-native applications, which are often deployed in distributed environments. Moreover, antivirus software is often resource-intensive and can slow down the performance of cloud-native applications if scalability is not properly considered. This is particularly problematic for cloud-native applications that require high performance, such as video streaming and machine learning.

Adding an in-built antivirus capabilities to MinIO would allow users to protect their data from malware without having to rely on third-party tools. This would also allow users to run antivirus software in a cloud-native with scalability and performance in mind. In this report we will discuss different approaches that can be taken to embed an antivirus engine into MinIO and evaluate their performance and scalability.

ClamAV

Open-source antivirus software is a popular choice for cloud-native and distributed applications. This is because open-source software is often free, which makes it a cost-effective option for users. Additionally, open-source software is often more customizable and flexible than proprietary software, which makes it easier to integrate into cloud-native applications. Open-source software is also often more secure than proprietary software, as it is often reviewed by a large community of developers and users.

One of the most popular open-source antivirus software is ClamAV. It is written in C and is available under the GNU General Public License (**?**). ClamAV is designed to be lightweight and fast, with a focus on distributed and scalable scanning. ClamAV uses a combination of signature-based and heuristic detection methods as mentioned earlier with a reported accuracy of XXXXX. It uses an open-source database of known malicious files and behaviors which is updated regularly by the ClamAV community.

For prototype implementation and benchmarking of the system architecture, we will use ClamAV as our antivirus engine. This can then be replaced or upgraded with a more advanced antivirus engine that utilises AI.

3. Specification and Design

Specification

Architecture

Design 1

Design 2

Design 3

Design 4

Optimal Design

Packaging

4. Implementation

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].

MinIO GitHub Repo. 2023. Minio github repository, Available at: https://github.com/minio/minio. [Accessed: 15 Mar 2023].

Ot, A. 2022. The object storage market, Available at: https://www.enterprisestorageforum.com/cloud/cloud-object-storage-market/. [Accessed: 17 Mar 2023].