Fortress: Secure De-duplicated Multi-Cloud Storage

Ghassan Karame, Wenting Li

Abstract

NLE's Secure De-duplicated Multi-Cloud Storage is the future of primary storage. This solution combines the use of multiple public cloud storage services with fast local access to cached data, data deduplication, and enhanced security and reliability at very low costs. By doing so, this solution is ideal for enterprise customers as well as governments and large telco-operators. Namely, our solution focuses on security and storage-efficiency, ensuring that the data is always available, is repaired in case of any partial data loss, and is protected from the strongest of adversaries, without hampering performance nor usability. As a result, our solution allows us to extend the range of current datacenter services and features, without incurring high infrastructure costs and at competitive margins.

Keywords

cloud storage, confidentiality, storage efficiency, de-duplication, retrievability, data loss detection

1. Introduction

Cloud services for storage, computing services, and collaboration platforms are becoming more relevant, important and pervasive. They offer a tremendous economic benefit to companies, private individuals, and public organizations. According to analysts, the amount of data storage is sharply increasing by approximately 50% per year; most of the increase in data storage corresponds to the storage of unstructured and archival data.

However, cloud services also introduce new security threats with respect to the confidentiality and integrity of their outsourced data. In fact, customers of cloud services lose control over their data and how data is processed or stored. This has been identified as the main obstacle for users to adopt cloud services.

This problem has been highlighted by the outbreak of the PRISM revelations in 2013. It became clear that customers cannot rely on their providers to protect their data. In addition to the well-known insider attacks, even "honest" cloud providers were coerced by the authorities to install backdoors or reveal the keys used to encrypt their customers' data. Therefore, we argue that existing commodity storage services which simply encrypt the data[1)] while retaining the encryption keys are not enough to address the omnipresent customers' concerns.

Our solution, Fortress, addresses this problem and is unique in tackling cloud customers' concerns. Fortress prevents data loss, automatically repairs data, and ensures data confidentiality at very low storage costs and high performance. Fortress uses multiple public cloud storage services, effective data de-duplication, and a novel security and retrievability technology to offer the lowest storage costs combined with the highest security protection. In this respect, Fortress combines two main pillars: (a) a novel data confidentiality primitive in the cloud, and (b) a novel data integrity and retrievability mechanism. In summary, the key features of Fortress are:

- Fortress ensures data confidentiality even if key is leaked using multiple public cloud storage services.
- Fortress improves storage efficiency using effective data de-duplication technology.
- Fortress prevents data loss and automatically repairs data using novel retrievability validation.

- Fortress provides accountability to the retrievability validation operations.

A unique feature of Fortress is that it offers verifiable services. Customers of Fortress can, at any point in time, verify its operations and get a non-refutable cryptographic proof of the correctness of Fortress software. This feature protects against internal misconfiguration errors that might arise from an improper installation of Fortress within the customer premises.

By doing so, Fortress provides security guarantees that cannot be offered by any other solutions in the market – including those costly solutions advertising in-house secure datacenters.

Fortress is therefore ideal for enterprise customers and for governments that are interested in providing the highest level of security for stored data while reducing storage costs by two orders of magnitude compared to any local secure storage solution.

In the following sections, we will describe in details the technology pillars in Fortress that give us the above security features. We also discuss in Section 3 the general deployment model of Fortress before concluding in Section 4.

2. Two Technology Pillars

Fortress makes use of two main security pillars to ensure confidentiality and retrievability in the strongest adversarial models. The first pillar ensures that sensitive data stored in existing clouds through our platform will remain highly protected even if the encryption keys are leaked[2]. Moreover, by leveraging a novel data de-duplication technology[3], Fortress is able to provide a secure storage service with higher degrees of scalability and at the lowest possible cost (**Fig. 1**). On the other hand, the second pillar guarantees that the customer data is always retrievable without any modifications in spite of malicious cloud administrators.

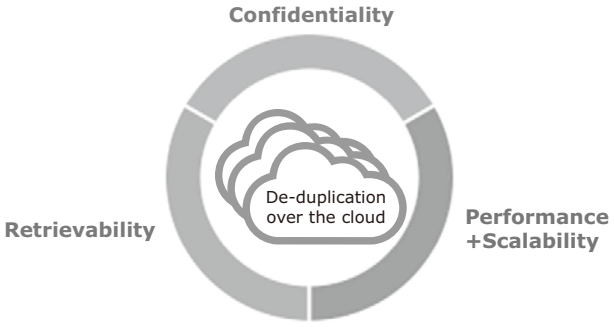

Fig. 1 Security and performance guarantees of cloud storage in Fortress.

We now start by discussing the two pillars of Fortress in details.

2.1 Pillar 1: Secure De-duplicated Encryption

Fortress leverages an NEC unique technology to provide data confidentiality against an adversary who knows the encryption key and can compromise a large fraction of the storage servers. Fortress is the first solution in the market that tolerates complete key leakage. Recent events[4] have shown that adversaries can acquire the keys either by exploiting flaws or backdoors in the key-generation software, or by compromising the devices that store the keys, both on the user-side and in the cloud. The latter can be caused by a careless administrator/user using a weak password or vulnerabilities in the system, such as misconfiguration errors.

2.1.1 Encryption with Multiple Clouds

Our technology adopts a special encryption primitive that combines the usage of multiple public clouds, with a novel all-or-nothing transform. Our technology guarantees no leakage of the customer data even if the attacker compromises the encryption keys and a large portion of the customers' data[2].

We contrast our technology to standard encryption primitives such as AES, which can also be combined with multiple clouds, but – unlike Fortress – will result in considerable data leakage if the adversary has acquired the encryption key. Namely, in standard symmetric encryption, if the adversary manages to acquire the encryption key and compromises a cloud provider (e.g., an insider or hacker), the adversary will be able to decrypt all data stored on the compromised cloud provider.

In contrast, Fortress uses a novel encryption primitive which ensures that the adversary must have access to all encrypted data and all keying materials in order to learn any bit of the message. In other words, even if the adversary manages to acquire the key and access the data fragment in one cloud, she is not able to acquire any bit of the data.

Fortress builds upon existing standardized block ciphers (i.e., AES), and only incurs 3% performance deterioration when compared to existing encryption primitives.

2.1.2 Encryption with De-duplication

Many cloud storage providers such as Dropbox[5] utilize de-duplication techniques to improve their storage efficiency. Namely, if the service identifies that the uploaded content has been already stored in the cloud (e.g., by

another user), data de-duplication is enforced to minimize the cost of the storage for duplicated data. Recent studies show that deduplication can result in up to 50% storage savings in standard file systems, and by up to 90% in backup applications. Such approach is, however, difficult to be applied on encrypted data as different users tend to encrypt their data using different keys thus deriving identical data in the storage.

Fortress uses a novel secure data deduplication technology[3] that enables storage optimization for encrypted content. Fortress leverages an oblivious server-assisted key generation protocol based on blinded BLS signatures in order to allow different (possibly untrusted) clients to acquire the same encryption key for the same data content.

2.2 Pillar 2: Secure Auditing of Data

None of the current cloud storage services accept liability for data loss in their Service Level Agreements (SLAs). Existing services only guarantee service availability – in spite of all the advertised cloud security and dependability solutions.

Fortress is the first solution in the market that provides its customers with a security SLA to account for data loss. Fortress provably verifies the availability and integrity of the data stored by its customers in the cloud[6]. While the main barriers of wide adoption of the cloud lie in the lack of customer trust and in the high costs of deploying security measures in cloud infrastructures, Fortress bridges these gaps and provides customers with the assurance that their files' status is constantly monitored.

If Fortress detects any partial data loss, it will immediately initiate correction strategies and repair the damaged content. Meanwhile, any external auditor or user can receive a proof that their data is still present in its entirety. This reduces the liability on the provider and protects against internal misconfiguration errors, and careless administrators, among others.

Fig. 2 depicts the data validation process in Fortress. Here, Fortress cryptographically challenges each utilized cloud provider in order to obtain provable and unforgeable evidence that the data is intact in their storage. Once data loss/corruption is detected (i.e., validation fails), Fortress will automatically repair the data using content stores at the remaining clouds.

In order to detect data loss in due time, the validation process needs to be efficient enough so that it can be frequently scheduled. In this respect, we note that Fortress's verification is extremely performant: it can cryptographically monitor the status of Terabytes of data within few seconds.

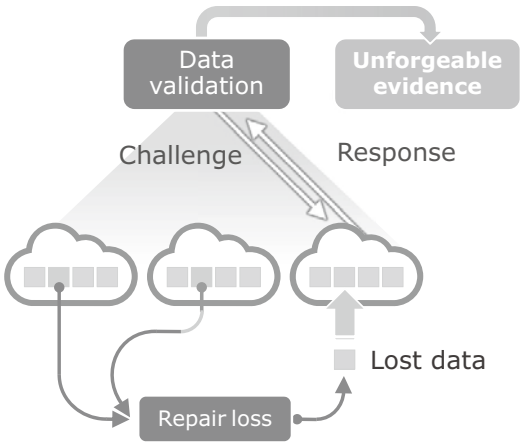

Fig. 2 Automatic data repair if loss is detected after a data validation.

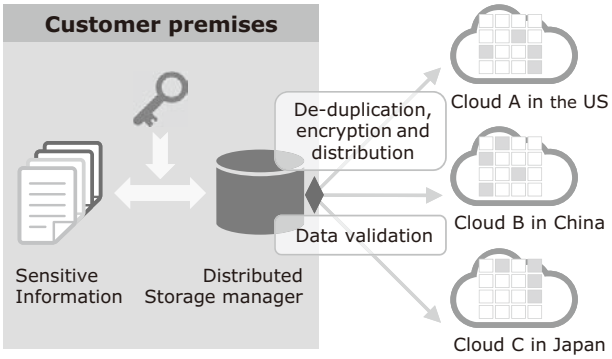

Fig. 3 Process flow of data processed by Fortress.

2.3 Summary

To summarize the two main technology aspects, **Fig. 3** shows how Fortress (the "Distributed Storage Manager") handles the customer sensitive information. On the one hand, when data is first requested to be stored, Fortress encrypts and distributes the data fragments among multiple clouds, and de-duplicates content if necessary. On the other hand, for all the stored data, Fortress periodically validates the retrievability of the data and repairs the data copy if needed.

3. Deployment Model

Fortress is a software layer that is typically installed on commodity servers placed within the customer premises. It interfaces with existing cloud providers, such as Amazon, OneDrive, Box.com, and Dropbox, using standard APIs without any modifications. Fortress can also support clouds that use the WebDav standard.

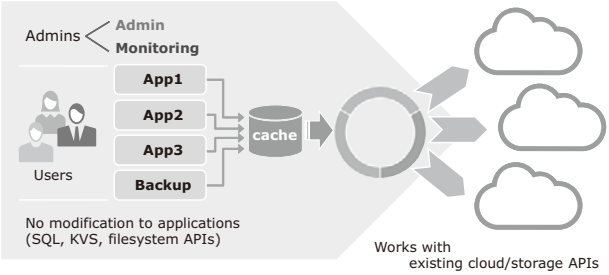

Fig. 4 Deployment model of Fortress.

As shown in **Fig. 4**, Fortress intercepts all calls made by the customer's applications to store data on disk, and redirects them to the cloud. All files sent to Fortress will be processed, and appropriately encrypted before being stored onto the cloud. Fortress's technology relies on standardized encryption primitives, such as AES and RSA.

Fortress interfaces with applications using SQL, Key Value Store (KVS) interface, as well as standard filesystem API. Fortress can also be extended to support additional application interfaces depending on the customer's request.

Due to its superior performance, Fortress is suitable not only for commodity servers, but can also be integrated within the software stack of Internet of Things (IoT) devices to securely extend the storage reach of these devices to the cloud.

4. Concluding Remarks

Fortress is a secure and performant cloud storage solutions for enterprise customers, large telco-operators, and governments. Fortress provides enhanced data confidentiality, efficient storage solutions, and verifiable service for data loss detection and data repair.

Fortress assures customers that the sensitive data they store onto existing clouds through our platform will remain highly protected, even if the encryption keys are leaked to a powerful adversary.

Fortress guarantees the customers that their data are always retrievable without any modifications, even when facing malicious cloud administrators. Fortress also automatically repairs content in case of any partial data loss.

By doing so, Fortress provides security guarantees that cannot be offered by any other solutions in the market – including those costly solutions advertising in-house secure datacenters.

Reference

1) B. Ken and H. Ryan, "Encrypting Data at Rest," White Paper of amazon web services, 2014.
2) K. Ghassan O., S. Claudio and L. Krzysztof, "Securing Cloud Data in the New Attacker Model," 2014.
3) A. Frederik, B. Jens-Matthias, K. Ghassan O. and Y. Franck, "Transparent Data Deduplication in the Cloud," in CCS, 2015.
4) Wikipedia, "Edward Snowden," [Online]. Available: http://en.wikipedia.org/wiki/Edward_Snowden#Disclosure
5) "Dropbox," [Online]. Available: www.dropbox.com.
6) A. Frederik, B. Jens-matthias, K. Ghassan O. and L. Zongren, "Outsourced Proofs of Retrievability," in CCS, 2014.

Authors' Profiles

Ghassan Karame
Senior Researcher
NEC Laboratories Europe
NEC Europe Ltd.

Wenting Li
Research Scientist
NEC Laboratories Europe
NEC Europe Ltd.

Information about the NEC Technical Journal

Thank you for reading the paper.
If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese English

Vol.10 No.3 Special Issue on Telecom Carrier Solutions for New Value Creation

Vol.10 No.3
July 2016

Special Issue TOP

General Paper

NEC Information