Towards Trustworthy Cloud Storage

Dan Dobre, Joao Girao, Ghassan Karame

Abstract

While public clouds are widely used for flexible deployment of online services such as video on demand, email, file sharing, etc., most enterprises still shy away from outsourcing sensitive data to public clouds because of security issues associated with storing data within a potentially untrusted cloud provider.
In this paper, we explore the solution space for enhancing the robustness and security of existing clouds. More specifically, we describe a set of cutting edge technologies that guarantee the security provided by the cloud infrastructure, the availability of its services, the privacy of its users, and the confidentiality of the stored data. Finally, we portray a "service image" that depicts the integration of these various technologies into one single multi-purpose cloud service.

Keywords

cloud attacker models, secure and dependable clouds, cloud service image

1. Introduction

The success of cloud computing is driven by the tremendous economic benefit offered to companies, private citizens and public administration to deploy, provision and use cloud services in a cost effective manner. The cloud is gaining increasing importance and applicability in numerous application domains. By shaping the way we access, store, and compute on data, the cloud offers a largely profitable business. Forrester projected that providing access to emails, and social networks will solely contribute towards a multi-billion dollar business by 2020.

Clearly, the larger are the utility and extent of the cloud and its services, the greater is the advantage in exploiting, abusing and attacking its components. As a matter of fact, Dropbox claims improper access to accounts using stolen passwords led to a spam attack, and Amazon and Gmail outages have left behind millions of disappointed customers not only because of access to their services but also those services (even third party) that depended on their cloud infrastructure.

The literature contains a number of similar incidents that threaten the availability of cloud services, the security of the underlying cloud infrastructure, and the privacy of the cloud customers and users. As a matter of fact, a very recent study by the Centre for European Policy Studies[1] concludes that security is not sufficiently addressed by current cloud technology and strongly underlines that "the challenge of privacy in a cloud context is underestimated, if not ignored."

In this paper, we explore the solution space for enhancing the robustness and security of existing clouds. More specifically, we categorize a multitude of cutting edge techniques according to realistic cloud attacker models. Finally, we portray a "service image" that depicts the integration of these various technologies into one single multi-purpose cloud service.

The remainder of this paper is organized as follows. In Section 2, we classify the various attacker models. In Section 3, we outline a taxonomy of solutions that enhance the security and dependability of data stored in the cloud. In Section 4, we present a novel service image that combines the described technologies in a single framework.

2. Threat Model

In this section, we outline several threats that are typically considered in the cloud setting. As we will see in Section 3, the strength of the adversarial model shapes the solution space, and impacts the design and complexity of the different approaches in terms of resource efficiency and performance.

2.1 Non-Responsive Cloud

Despite considerable investments made by cloud providers into the reliability of their infrastructure, many of them have suffered from outages in the recent past[3], whereupon the availability of the data has been compromised. Even in the absence of outages by the cloud storage service itself, users' data can become unavailable due to connectivity problems (e.g. a network partition) or due to unexpected contractual changes to the disadvantage of the user (provider lock-in). Many modern applications relying on the cloud storage services are delay intolerant, and data unavailability can have disastrous consequences for them. Outages of the data storage service, as well as network outages are typically captured by the Non-Responsive Crash model, in which individual storage services may fail by crashing and the links may fail to deliver messages in a timely fashion. Storage protocols designed to tolerate such (mis-)behavior are termed crash fault-tolerant.

2.2 Honest but Curious Cloud

The community features a number of incidents where sensitive client data has been leaked outside the cloud. This was caused either by (i) dishonest cloud and IT operators that can potentially leak sensitive data to high-bidding third parties, and/or (ii) due to an intrusion within the cloud premises (e.g., a hacker or malware).

Here, although the cloud itself might not be malicious, it cannot be always trusted to maintain the confidentiality of the client data. This is captured by the "honest but curious" threat model, which builds upon the prior "crash" model.

2.3 Rational Cloud

Unlike the honest but curious model, rational clouds are untrusted clouds that aim at maximizing their benefit in the system according to a predefined utility function. Examples include cloud operators storing data (i) at lower redundancy levels than advertised, (ii) in different locations (i.e., with cheaper cost of storage) than requested. Moreover, rational clouds seek to log data access patterns and user preferences in order to profile the cloud users e.g., for the purpose of targeted advertisement or on the behalf of interested third parties. Clearly, this is a severe departure from the "honest" model (i.e., crash and honest but curious models) since the cloud might have considerable incentives to deviate from the specified protocols.

2.4 Byzantine Cloud

As far as we are aware, Byzantine clouds constitute the strongest threat model in which the cloud may act arbitrarily malicious (even to its economic disadvantage). This Byzantine behavior can result from both internal and external exploits of the cloud infrastructure and services. Intuitively, Byzantine threat models capture all types of misbehavior, including that of a rational cloud. Examples of Byzantine behavior include a potential compromise of data integrity, and confidentiality at large scale. One of the main challenges in Byzantine clouds arises from bounding the number of Byzantine components by a threshold over time.

3. Enabling a Secure and Dependable Cloud

In this Section, we describe a taxonomy of solutions to the threat model outlined in Section 2. We describe the integration of these individual technologies in a unified framework in Section 4.

3.1 Non-Responsive Cloud

To cope with outages of individual data storage services, a number of solutions exist that rely on a collection of storage clouds (**Fig. 1**). One of the main challenges is to guarantee high availability and strong consistency despite failures and concurrent access to shared data. The basic functionality exposed to clients is the read/write storage interface, which constitutes the core of Key-Value Store (KVS) APIs - the de-facto standard of modern cloud storage offerings.

There are a number of recent works that aim at implementing a highly available and strongly consistent cloud storage substrate by replicating or striping data (in the vein of RAID) onto multiple storage clouds. These works rely on the KVS API exported by the individual clouds, typically consisting of basic operations *op* $\in$ *{Put, Get, Delete, List}*, and the more advanced *Conditional <op>*. Unlike the basic operations, conditional operations are not supported by every cloud storage provider. However, given that conditional operations enable transactional access to the strongly consistent data, they have

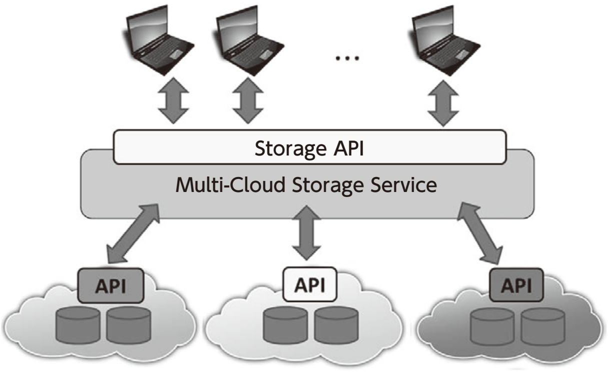


Fig. 1 Sketch of a multi-cloud storage service.

become an integral part of many storage offerings, e.g. Windows Azure Storage, Yahoo! Peanuts, Amazon DynamoDB, to name a few.

In a recent work, Libdeh et al.[14] proposes RACS, a multi-cloud storage system in which the data is striped on a collection of storage clouds, each of which exports a basic KVS API. RACS relies on an external coordination service (i.e. a proxy) which mediates the entire interaction between clients and the storage cloud. Basescu et al.[15] present ICStore, IBM's implementation of a multi-cloud data sharing service. The access to the data is coordinated by the clients in a decentralized fashion without inter-client communication. In contrast to RACS, in which all requests are channeled to a (logically) centralized proxy, ICStore's ability to scale is only limited by the scalability of the underlying storage clouds. Unlike RACS, that stripes data across clouds to save storage costs, ICStore relies on data replication. Furthermore, ICStore features strong consistency (i.e., reads always "see" the most recent update) and high availability despite network outages and crash failures by a minority of clouds, and any number of clients. In the best case, for each data item stored, ICStore incurs a storage overhead of 2 data items per cloud, which is twice as much as required by a single cloud solution. In the worst case, the space overhead associated with storing each individual data item is proportional to the number of clients that may update the data item.

3.1.1 NEC Technology

NEC has developed patented technology aimed at reducing the space usage incurred by existing multi-cloud storage solutions. Specifically, Chockler, Dobre and Shraer[16] propose MCStore, a multi-cloud storage service that leverages conditional Put operations at individual clouds to achieve constant space overhead. Specifically, for each data item, MCStore requires storing a single data item per cloud, regardless of the number of clients. In addition, the latency of MCStore, measured as the number of cloud accesses per operation, adapts to the number of concurrent clients. As a result, in the absence of concurrency (which is the common case), MCStore achieves optimal latency of 2 cloud accesses per operation.

3.2 Honest but Curious Cloud

The literature features a number of solutions to protect against an honest but curious cloud. At the core of these solutions is efficient data encryption in such a way that encryption key cannot be accessed by any single cloud operator. The main challenge that arises here is key management.

There are three main approaches to manage encryption keys to resist an honest but curious cloud. The simplest approach is to share the encryption key among the users of the cloud, and to never expose that key to the cloud. In this way, the confidentiality of the data in the cloud can be ensured since only users can decrypt the data. However, this approach requires a cumbersome key management process to share the key among the cloud users.

Another approach would be to keep the en-cryption key in the cloud, but secret-share[3] it among the cloud servers, so that any t servers cannot acquire the key. To acquire the key, t+1 servers need to collude and reconstruct the secret-shared key. Similarly, to decrypt an encrypted data object, users have to acquire t+1 shares of the key in order to be able to reconstruct the key and decrypt content.

One of the most efficient approaches to resist to an honest but curious cloud model would be to rely on a transformation called All or Nothing Transformation (AONT)[4] [5] [6] AONT are typically a keyless transform; as such, these transforms do not incur any overhead in terms of key management. AONTs mainly consists of dispersing the encrypted data across different cloud servers and ensures that the data cannot be retrieved by any single server, unless this server has access to all the data that is stored on all the servers, hence the name, All or Nothing. AONTs can be practically constructed by relying on block ciphers and by embedding the encryption key within the ciphertext itself. An exemplary construct for an AONT scheme can be found in Reference[5].

3.3 Rational Cloud

The literature features a number of solutions that allow a cloud to prove to remote entities that the cloud is behaving correctly.

A number of solutions aim at proving to remote entities that the cloud possesses/stores data. These are often referred to in the literature by Proofs of Retrievability (PoRs)[7] and Proofs of Data Possession (PDPs)[8] [9] protocols. Both PoRs and PDPs are challenge-response protocols that enable the remote verification of file possession while minimizing the incurred communication complexity. The basic idea here is to embed cryptographic tags within the data. Since, the client possesses the key, he/she can verify the authenticity of the tags.

Proofs of Location (PoLs)[10] combine POR schemes with geolocation systems and aim at verifying that a given file is correctly stored at a given location.

Similarly, Proofs of Redundancy can be used to ensure that data is replicated among several servers[11]. Proofs of Redundancy can be achieved by leveraging the use of PoLs and verifying that the same file is stored across multiple geolocation. Bowers et al.[14] also presented a scheme that enables the construction of Proofs of Redundancy for data that is stored in a single geolocation. The literature also contains other proposals e.g., that verify that content is stored in encrypted in the cloud[3].

3.4 Byzantine Cloud

Byzantine-resilient distributed storage has attracted considerable research attention, due to the appealing promise of protecting the user from arbitrary data corruption, compromised availability and consistency violations. HAIL[17] is a distributed cryptographic storage system that extends the use of PORs to multiple servers to ensure integrity protection and availability (retrievability) of files dispersed across several storage clouds. HAIL assumes a Byzantine cloud model and considers a mobile adversary and a single client interacting with the storage in a synchronous fashion. IRIS[11] is a PoR-based distributed file system designed with enterprise users in mind that stores data in the cloud and is resilient against potentially Byzantine service providers. Although IRIS is designed to cope with multiple clients, all operations are pre-serialized by a logically centralized, trusted portal which acts as a fault-free gateway for communication with untrusted clouds. Bessani et al.[20] propose DepSky, a storage system for confidential data, relying on erasure coding, digital signatures and secret sharing of encryption keys. DepSky supports multiple writers through an external locking mechanism that determines a unique writer, and which requires clients to communicate with each other.

3.4.1 NEC Technology

NEC has developed patented technology aimed at increasing the performance of existing storage solutions. Precisely, Dobre et al.[20] present PoWerStore, a highly available and strongly consistent distributed storage protocol based on lightweight crypto (i.e., cryptographic hashes and message authentication codes). Central to PoWerStore is the concept of “Proofs of Writing” (PoW), a novel data storage technique inspired by commitment schemes. PoW rely on a 2-round write procedure, in which the first round writes the actual data and the second round only serves to “prove” the occurrence of the first round. PoW enable efficient implementations of strongly consistent storage across Byzantine clouds through metadata write-backs and low latency reads. Unlike HAIL, PoWerStore assumes a static adversary, yet assumes a distributed client setting in which clients share data in an asynchronous fashion. In contrast to DepSky, PoWerStore relies on the highly available distributed PoW technique, obviating the need for direct communication among clients.

4. Unified Service Image

The cloud attacker models presented in Section 2 provide us with a set of system requirements to address the cloud storage needs of future customers. Although market leader Amazon provides a simple, pay as you go solution, the market left untapped is looking for a more flexible and secure cloud system. The “one size fits all” competition in price is neglecting customers who would pay more to address their security needs.

Key to channelling these new customers is addressing the system requirements at a service level. A customer must not only believe that its data is properly secured but understand what impact different options have in the service and business.

Based on the attacker models and technologies described in this paper, we focus the customers’ attention in three distinct groups:

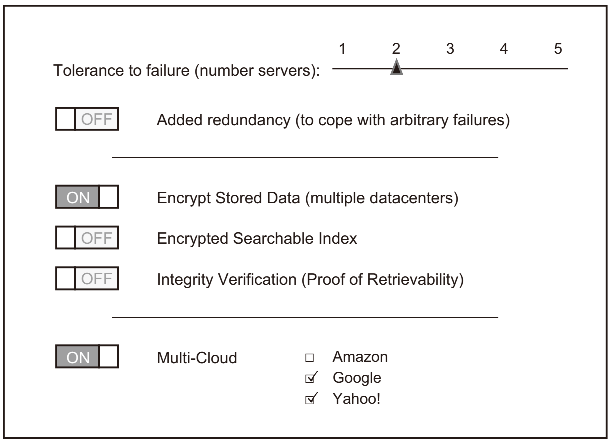
- **Tolerance to failure:**
  how many servers can be affected and what type of failures can the system endure without affecting service.
- **Security properties:**
  what type of security parameters should be observed by the system: here we focus on confidentiality and integrity.
- **Portability and Multi-domain:**
  using multiple cloud providers can reduce the risk of information leakage, increase reliability and facilitate data portability.

Fig. 2 depicts a mock-up of a service image as it could appear to the customer. Depending on the customer needs, it is simple to understand what extra security and dependability services the customer may need. Price of the service is adapted to service guarantees. When no extra guarantees are requested, the storage service can compete on price.

Using this simple interface, the customer can choose which attacker model to address by selecting security features. The customer does not need to fully understand the attacker model or the technology behind the features. We can balance the security features of the storage service with the needs of the customer and the price he is willing to pay.

5. Conclusion

Although the cloud offers many performance advantages



Tolerance to failure (number servers): 1 2 3 4 5

OFF Added redundancy (to cope with arbitrary failures)

ON Encrypt Stored Data (multiple datacenters)

OFF Encrypted Searchable Index

OFF Integrity Verification (Proof of Retrievability)

ON Multi-Cloud
□ Amazon
☑ Google
☑ Yahoo!

Fig. 2 Service image.

with respect to both storage and computation, it unfortunately raises many security and privacy exposures. One of the main reasons why businesses and governments still do not store their sensitive data in the cloud is the lack of trust in the service provider. In this paper, we described a number of techniques addressing various security, privacy and dependability challenges in the face of potentially malicious cloud providers and how these techniques can be presented to a customer.

At NEC we focus on the performance and functionality covered by these techniques. We have developed space-efficient storage in multi-cloud environments using existing APIs and the most efficient byzantine fault-tolerant distributed storage protocol. We have also developed a highly efficient AONT, and con-tinue to find more efficient solutions to problems such as searchable encryption, access control and verifiable policy enforcement to strengthen our security and dependability portfolio for cloud storage services.

* Dropbox is a trademark or a registered trademark of Dropbox, Inc. in the US.

* Amazon is a trademark of Amazon.com, Inc. and/or its affiliates.

* Gmail is a trademark or a registered trademark of Google Inc.

* Windows and Windows Azure are registered trademarks of Microsoft Corporation in the US and other countries.

* Yahoo! is a trademark or a registered trademark of Yahoo! Inc. in the US.

* IBM is a registered trademark of International Business Machines Corporation in the US and other countries.

Authors' Profiles

Dan Dobre
Research Scientist
NEC Laboratories Europe
NEC Europe Ltd.

Joao Girao
Manager
NEC Laboratories Europe
NEC Europe Ltd.

Ghassan Karame
Research Scientist
NEC Laboratories Europe
NEC Europe Ltd.

Reference

1) Didier Bigo,et al. : Policy Department C: Citizens' Rightsand Constitutional Affairs
http://www.europarl.europa.eu/committees/en/studies-down-load.html?languageDocument=EN&file=79050
2) 2009 Sidekick data loss
http://en.wikipedia.org/wiki/2009_Sidekick_data_loss
3) A. Shamir: How to Share a Secret? , Communications of the ACM, pages 612–613, 1979.
4) V. Boyko : On the Security Properties of OAEP as an Allornothing Transform, CRYPTO' 99, pages 503–518, 1999.
5) R. Rivest : All-or-Nothing Encryption and The Package Transform, FSE '97 Proceedings of Fast Software Encryption, pages 210–218, 1997.
6) D. R. Stinson: Something About All or Nothing (Transforms), In Designs, Codes and Cryptography, pages 133–138, 2001
7) K. Bowers, et al. : Hail: a high-availability and integrity layer for cloud storage, CCS, pages 187-198. 2009.
8) G. Ateniese, et al. : Provable data possession at untrusted stores, CCS, pages 598-609. 2007.
9) Emil Stefanov, et al. : Iris: a scalable cloud file system with efficient integrity checks, ACSAC, pages 229-238, 2012.
10) Gaven J. Watson, et al. : LoSt: Location Based Storage, Proceedings of CCSW 2012.
11) Karyn Benson et al. : Do You Know Where Your Cloud Files Are?, Proceedings of CCSW 2011.
12) K. D. Bowers, et al. : How to tell if your cloud files are vulnerable to drive crashes, CCS 2011.
13) van Dijk, et al. : Hourglass schemes: how to prove that cloud files are encrypted, CCS 2012.
14) H. Abu-Libdeh, et al. : RACS: a case for cloud storage diversity, SoCC, pages 229–240, 2010.
15) Cristina Basescu, et al. : Robust data sharing with keyvalue stores, DSN, pages 1–12, 2012.
16) Gregory Chockler, et al. : Consistency and Complexity Tradeoffs for Highly-Available Multi-Cloud Store,
http://people.csail.mit.edu/grishac/mcstore.pdf
17) Kevin D. Bowers, et al. : Hail: a high-availability and integrity layer for cloud storage, CCS, pages 187-198, 2009.
18) Dan Dobre, et al. : Proofs of Writing for Efficient and Robust Storage,
http://arxiv.org/abs/1212.3555
19) Emil Stefanov, et al. : Iris: a scalable cloud file system with efficient integrity checks, In ACSAC, pages 229-238, 2012.
20) A. Bessani, et al. : Dependable and secure storage in a cloud-of-clouds, EuroSys, 2011.

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.8 No.1 Solving Social Issues Through Business Activities

Vol.8 No.1

September, 2013

Special Issue TOP