Hybrid Cloud in data deduplication

Sreeba K A,
Asst. Prof. on contract,
Dept. Of Computer Science,
Carmel College, Mala

Abstract: **Information deduplication is one of important Information compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive Information while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect Information security, it makes an attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself.**

INTRODUCTION

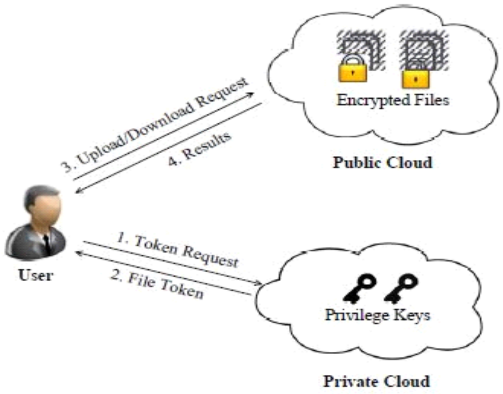
As cloud computing becomes prevalent an increasing amount of Information is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored Information. One critical challenge of cloud storage services is the management of the ever increasing volume of data. To make data management scalable in cloud computing, deduplication has been a well known technique for eliminating duplicate copies repeating Information in storage. The technique is used to improve the storage utilization. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant Information by keeping only one physical copy and referring other redundant data to that copy.Traditional encryption requires different users to encrypt their Information with their own keys. Thus identical data copies of different users will lead to different cipher texts, making deduplication impossible. Here convergent encryption used .Here it encrypts or decrypts a data copy with a convergent key which is obtained by computing the cryptographic hash value of the content of the data copy. Here hybrid cloud approach used for safe and secure data storage which consisting of private cloud and public cloud. Private cloud is involved as a proxy to allow data owner or user to securely perform duplicate check with differential privileges. The encrypted Information stored in public cloud. The Information owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

WORKING OF DEDUPLIACTION

Hybrid cloud is an IT architecture that incorporates some degree of workload portability, orchestration, and management across 2 or more environments. Depending on whom you ask, those environments may need to include:

- At least 1 private cloud and at least 1 public cloud
- 2 or more private clouds
- 2 or more public clouds
- A bare-metal or virtual environment connected to at least 1 cloud—public or private

These varying requirements are an evolution from the earlier age of Cloud computing, where the differences between public clouds and private clouds were easily defined by location and ownership. But today's cloud types are far more complex, because location and ownership are abstract considerations.



Private cloud

A private cloud consists of computing resources used exclusively by one business or organisation. The private cloud can be physically located at your organisation's on-site datacenter or it can be hosted by a third-party service provider. But in a private cloud, the services and infrastructure are always maintained on a private network and the hardware and software are dedicated solely to your organisation. In this way, a private cloud can make it easier for an organisation to customise its resources to meet specific IT requirements. Private clouds are often used by government agencies, financial institutions, any other mid- to large-size organisations with business-critical operations seeking enhanced control over their environment.

Public cloud

Public clouds are the most common way of deploying cloud computing. The cloud resources (like servers and storage) are owned and operated by a third-party cloud service provider and delivered over the Internet. Microsoft Azure is an example of a public cloud. With a public cloud, all hardware, software and other supporting infrastructure is owned and managed by the cloud provider. In a public cloud, you share the same hardware, storage and network devices with other organisations or cloud "tenants."

Hybrid cloud

Often called "the best of both worlds," hybrid clouds combine on-premises infrastructure, or private clouds, with public clouds so organisations can reap the advantages of both. In a hybrid cloud, data and applications can move between private and public clouds for greater flexibility and more deployment options. For instance, you can use the public cloud for high-volume, lower-security needs such as web-based email and the private cloud (or other on-premises infrastructure) for sensitive, business-critical operations like financial reporting. In a hybrid cloud, "cloud bursting" is also an option. This is when an application or resource runs in the private cloud until there is a spike in demand (such as seasonal event like online shopping or tax filing), at which point the organisation can "burst through" to the public cloud to tap into additional computing resources.

Advantages

- Cost-effectiveness—with the ability to scale to the public cloud, you pay for extra computing power only when needed.
- Ease—transitioning to the cloud does not have to be overwhelming because you can migrate gradually—phasing in workloads over time.
- Control—your organisation can maintain a private infrastructure for sensitive assets.
- Flexibility—you can take advantage of additional resources in the public cloud when you need them.

Which one should you choose?

It's important to keep in mind when deciding whether to build a private or public cloud, to properly weigh the differences against each other. In most cases they can be thought of as advantages or disadvantages, depending on the usage required. If we'd like to store our backup data somewhere in the cloud, it's important to determine the sensitivity of said Information. For example, if we are storing confidential information such as credit card information or medical records we absolutely must store that data in a private cloud but when it comes to non-sensitive info, we can store it in a public cloud if it keeps costs down considerably.

Then, there's always a choice whether to integrate public or private cloud into our everyday operations. Again, there are advantages and disadvantages that need to be taken into consideration. Whether to compromise the security and keep expenses down or pay a little extra for additional layer of security is a choice you'll sooner or later have to make; but you should always consider what's best for your business and move forward from there.

REFERENCES

[1] OpenSSLProject. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296 312, 2013.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for Identity-based identification and signature schemes. J. Cryptology, 22(1):161, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162177, 2002.

[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617624, 2002.

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

[10] GNULibmicrohttpd. http://www.gnu.org/ software/ libmicrohttpd/.

[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491500. ACM, 2011.

[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou.