ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Download Speed and Reliability Challenges of Network Coded Storage Systems

Manasa Jonnalagadda
B.Tech, CSE, III-I,
G.Narayanamma Institute of Technology & Science.

ABSTRACT:

Although cloud systems provide a reliable and flexible storage solution, the use of a single cloud service constitutes a single point of failure, which can compromise data availability, download speed, and security. To address these challenges, we advocate for the use of multiple cloud storage providers simultaneously using network coding as the key enabling technology. Our goal is to study two challenges of network coded storage systems. First, the efficient update of the number of coded fragments per cloud in a system aggregating multiple clouds in order to boost the download speed of files. We developed a novel scheme using recoding with limited packets to trade-off storage space, reliability, and data retrieval speed. Implementation and measurements with commercial cloud providers show that up to 9x less network use is needed compared to other network coding schemes, while maintaining similar download speeds and reliability. Second, the ability to update coded fragments from a linear erasure code when the original file is modified. We exploit code structure to provide efficient representations of the evolution of the file. Evaluations using file changes on software library repositories show that a five-order of magnitude reduction in network and storage use is possible compared to state-of-the-art.

INTRODUCTION:

Cloud storage is widely adopted as it offers a costeffective solution to storing enterprise data, with the advantages of increased reliability as well as arguably decreased technical complexity and business agility compared to on-site, personalized storage solutions. Its adoption among end users is growing as well thanks in part to the free storage offered by major IT players, e.g., Amazon, Google, Microsoft, Apple, and specialized cloud storage companies, e.g., Dropbox, Box, SugarSync. A limitation of these solutions is that users are typically tied to a single cloud storage provider and, thus, dependent on the provider's offered reliability and availability of the data. The impact of an outage of a single provider can be significant.As pointed out in , another issue is data privacy. Even when the data is encrypted, it is only stored by a single provider, which can make it particularly vulnerable to attacks or even disclosure to governmental bodies of a foreign country. To address these limitations, network coding constitutes a particularly interesting technology. Although it was originally conceived for enhancing communication networks, it has shown significant potential in distributed storage applications, e.g., One key behind the impact of network coding lies in the fact that not only sources can generate coded packets. In fact, intermediate nodes in the network can generate coded packets from data available in their buffers without having all packets needed to decode, in a process known as recoding . Another important breakthrough came from the creation of random linear network coding (RLNC) . RLNC is a distributed approach that allows nodes to generate coded packets by creating linear combinations of existing coded packets using random coefficients. RLNC requires no coordination between nodes and preserves performance and code properties.

EXISTING SYSTEM:

In our previous work , we have shown using measurements performed with 5 cloud providers that a network coded multi-cloud storage solution outperforms a single cloud approach. We have also shown that the aggregation of clouds is much more effective when using RLNC instead of replication repetition-based codes employed in datacenters. We have conducted a comparative study on the effectiveness of RLNC in maintaining data integrity in storage and bandwidth constrained dynamic scenarios. We have found RLNC to surpass repetition-based and Reed-Solomon codes used in state of the art systems in both centrally controlled scenarios such as datacenters as well as fully decentralized scenarios including peer-to-peer storage systems. We therefore consider an RLNC-based distributed storage system as state of the art. As we have not encountered any commercial systems that employ or research that describes adaptation in the data distribution, we compare our proposed techniques with a non-adaptive approach.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

DISADVANTAGE:

» Although cloud systems provide a reliable and flexible storage solution, the use of a single cloud service constitutes a single point of failure, which can compromise data availability, download speed, and security.

PROPOSED SYSTEM:

Our work focuses on distributed storage solutions using RLNC. In this paper, we present a system that employs commercially available clouds to store files reliably. Given the similar challenges, our solutions can be adapted for use with other types of storage nodes as well. Our proposed system is comprised of a client application that uploads and downloads data to the storage nodes and handles all computations related to encoding, decoding and recoding. The storage nodes have no other functionality besides storing the data, which makes employing recoding techniques Those involve encoding on the nodes impossible. This is a limitation in several commercial clouds. We have built a model without this limitation and explored other types of recoding separately to this work.

ADVANTAGE:

Cloud storage is widely adopted as it offers a cost effective solution to storing enterprise data, with the advantages of increased reliability as well as arguably Decreased technical complexity and business agility compared to on-site personalized storage solutions.

IMPLEMENTATION:

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

IMPLEMENTATION MODULE DESCRIPTION:

After careful analysis the system has been identified to have the following modules:
1.Data Owner
2.Cloud

1.Data Owner:

In this module if a owner of data(File) have to store data on a cloud server, he/she should register their details first. These details are maintained in a Database. Then he has to upload the file in a file database. The file which are stored in a database are in an encrypted form. Authorized users can only decode it.

2. Cloud:

In this module CSP has to login first. Then only he can store the file in his cloud server. All file can only check the csp whether the csp is authorized csp or not.If its fake, wont allow the file to store in cloud server.

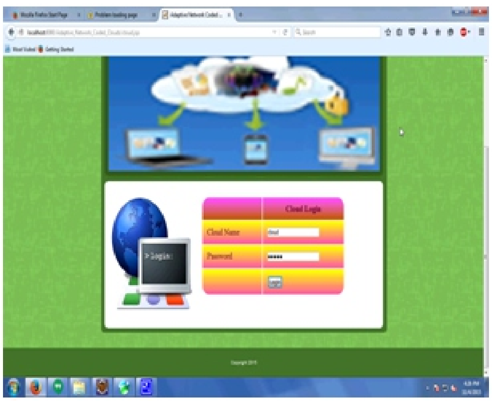
SCREEN SHOTS

Home:

File Upload:

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Cloud Login:



CONCLUSION:

In this paper we presented the benefits of using random linear network coding with multiple commercially available cloud storage providers to create a cloud of clouds. We decided on tackling this scenario from several perspectives and included two distinct contributions. First, we proposed a novel data distribution mechanism and showed that it can almost match on performance a more wasteful symmetric one. We proposed a novel technique which adapts the distribution of data to the state of the individual clouds. It minimizes the required bandwidth needed for the adaptation steps by doing dense recoding only for data that ensures recoverability and sparse recoding for performance enhancing data. We have shown with measurements that these two mechanisms greatly increase performance and result in significant savings on adaptation bandwidth. Second, we presented a detailed mechanism to support file updating and version control when using erasure correcting codes for storage applications, including any random linear network coded system. These ideas can be applied to standard data center systems as well as novel cloud storage and peer-to-peer distributed storage mechanisms that are keen on supporting various versions of the same file without high storage costs or for efficient updates without the inherent costs of uploading full files to all storage nodes. We showed that our proposed technique is a viable and extremely effective solution by applying it on real-world data taken from a Git repository.

REFERENCES:

[1] Extended version of the paper available online at http://tinyurl.com/3ke8ese.

[2] The Coding for Distributed Storage wiki http://tinyurl.com/storagecoding .

[3] J. Li, S. Yang, X. Wang, B. Li. “Tree-structured data regeneration in distributed storage systems with regenerating codes,” in Proc. IEEE Infocom 2010, April 2010.

[4] A. G. Dimakis, P. G. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” in IEEE Trans. on Inform. Theory, vol. 56, pp. 4539 – 4551, Sep. 2010.

[5] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, “A survey on network codes for distributed storage,” in IEEE Proceedings, vol. 99, pp. 476 – 489, Mar. 2011.

[6] A. Asterjadhi, E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi. "Towards network coding-based protocols for data broadcasting in wireless ad hoc networks." in IEEE Transactions on Wireless Commun., vol. 9, pp. 662 – 673, Feb. 2010.

[7] C. Gkantsidis, J. Miller, and P. Rodriguez, "Comprehensive view of a live network coding P2P system," in IMC, Association for Computing Machinery, Inc., Oct 2006.

[8] O. Khan, R. Burns, J. Plank, and C. Huang, "In search of I/O-optimal recovery from disk failures," to appear in Hot Storage 2011, 3rd Workshop on Hot Topics in Storage and File Systems, Portland, OR, Jun., 2011.

[9] H. Weatherspoon and J. D. Kubiatowicz, "Erasure coding vs. replication: a quantitative comparison," in Proc. IPTPS, 2002.

[10] M. Blaum, J. Brady, J. Bruck, and J. Menon, "EVENODD: An efficient scheme for tolerating double disk failures in raid architectures," in IEEE Transactions on Computers, 1995.