ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Survey on Beehive Erasure Codes for Distributed Storage Systems

Mugdha Adivarekar,Vina Lomte

M.E. Student, Dept. of Computer Engg, RMD Sinhgad School of Engineering, S. P.Pune University, Pune, India

Asst. Professor &HOD, Dept. of Computer Engg, RMD Sinhgad School of Engineering, S. P. Pune University, Maharashtra, India

ABSTRACT: To deal with the common security issues like data confidentiality and data robustness in cloud storage is very challenging. Encryption/Decryption technique is used for maintaining data confidentiality while data robustness has two concerns: service failure, and service corruption. In this paper Beehive erasure code in distributed network is majorly taken into consideration to find solution for data robustness of system. This system takes care of data confidentiality as it calculates lost data on failed server simultaneously using erasure codes. Beehive codes as compared to traditional Reed Solomon codes are more optimized for network transfer volume which slightly incurs more storage than Reed Solomon code but this storage overhead can be reduced by using MDS codes. MDS code is improvised version of Beehive code. Integrity checks can be added explicitly against storage server corruption.

KEYWORDS: Data confidentiality, Data robustness, Beehive code, Integrity check, Decentralized erasure code

I. Introduction

In distributed cloud storage network many problems are faced because ofsudden failures of servers.So we need construct a distributed storage system which will support data integrity,data confidentialityand making system robust enough to face service failure and service corruption.

Here we will focus on i) security attacks, ii) Security mechanisms and (iii) Service Continuity.

So to fulfil all these requirements this topic helps to implement secure distributed storagewhere system consists of storage servers (SS) at various locations.

There are following type of failures that can occur in distributed systems –

- Failure Caused by Human Fault
- Failure Caused by Bugs in Application
- Failure Caused by Service Provider Fault
- Failure Caused by Quality Failure
- Failure Caused by Increasing Demands
- Failure Caused by Security Reasons
- Failure Caused by Service Failure
- Failure Caused by Container Failure
- Failure Caused by Bad recovery policies

In all of the cases above it becomes very challenging to restore lost data on failed servers. For common human beings though it is not that important to save their data safely, for a businessman data safety matters a lot.

.

II. Related work

H.-Y. Lin and W.-G. Tzeng [6] first started research regarding Decentralized Erasure codeto enhance the privacy/confidentiality and robustness of the distributed storage network. Their research paper addresses the solution at low computation and storage cost. Decentralized erasure code is random linear code with sparse generator matrix.
Then H.-Y. Lin andW.-G. Tzeng [5] have initially proposed a threshold proxy re-encryptionscheme and integrate it with decentralized erasure code so that secure erasure code basedstorage network system is formulated.Re-encryption scheme allows encoding operations onencrypted data to provide data security. Also worked on retrieving file contents when theyare split and erasure code is added to them. Combining these divided part sequentiallywhile retrieving is one of the major challenges.
Shiuan-Tzuo Shen, Hsiao-Ying Lin, Wen-Guey Tzeng [2] have proposed to provide dataconfidentiality by means of adding integrity tags in data content before encryption. Thispaper is based on "A secure erasure code-based cloud storage system with secure data forwarding"as it gives prerequisites of the system i.e. data security.This is the base paper whichI have referred.
LaszloCzap,ChristinaFragouli,VinodPrabhakaran,SuhasDiggavi [3] elaborates more abouthow erasure code is generated and how it works during server failure in network so thatmissing component of file is retrieved easily by generating that missed component and alsoit subjects to Feedback of channel state on Eve and legitimate networks. Also if networkchannel introduce errors,then it would apply channel code to correct them. By leveragingerasures and feedback, secrecy rates can be increased.
Jian Liu,Kun Huang,Hong Rong,Huimei Wang and Ming Xian[4] contributes to present atechnique of efficient data forwarding scheme for erasure coded and encrypted cloud, whichenforces cloud to provide reliability, confidentiality and forwarding same data to another userwithout retrieving back.This paper proposed Reed Solomon erasure code scheme to generate beehive erasure codeswith less network transfer and they can reconstruct data on multiple storage servers simultaneously
.

A. *Existing Methodologies*:

- For providing maximum availability Distributed cloud storage provides backup servers which have replicas over different locations in world.
- These servers implement RAID to have data copies with maximum fault tolerance and dataavailability.
- Code based methods for making replicas and Error correction codes are now rarely seenbut they do exists in some scenario for archival and effective storage.
- Single storage server for whole file
- Array codes
- Non-MDS Codes

B. *Drawbacks*

1) Hidden Cost and Additional overheads of replicas
2) Distance multiplies risks
3) May leak stored data due to lack integrity proofs
4) Only one time encryption using general encryptionschemes
5) The user has to do most computation andcommunication traffic between the user and storagedevices is high.
6) The user has to manage his cryptographic keys.
7) If the user's device of storing the keys is lost orcompromised, the security is broken.
8) It is hard for storage servers to directly forward a user'smessages to another one.
9) The owner of the message has to retrieve, decode,decrypt and then forward them to another user.

C. Problem Analysis

To provide data robustnessstoring multiple copies of the original databrings expensive overhead to the storage system. This provides a trade-off between the storage size and thetolerance threshold of failure servers.

Therefore, distributed storage systems have been replacing replication with erasure codes, especially for cold or archival storage.

Once one block becomes unavailable, the missing datashould be fixed through a reconstruction operation andsaved at some other server.

A decentralized erasure Homomorphic integrity tags are also compatible with data storage, forwardingand repairing to check confidentiality. System consists of n storage servers, m keyservers and 4 phases:

1. Setup
2. Storage and Design of topology
3. Integrity check (Optional)
4. Reconstruction and Retrieval phase

Server A requests for integrity check for File M of identifierI_m to key server KS.KS queries random w servers for w encoded tuples tocalculate integrity checks.

III. Proposed System

A. *Design Considerations:*

System consists of 4 phases – Setup,Storage and Design topology,Integrity Checkand Reconstruction and Retrieval Phase.

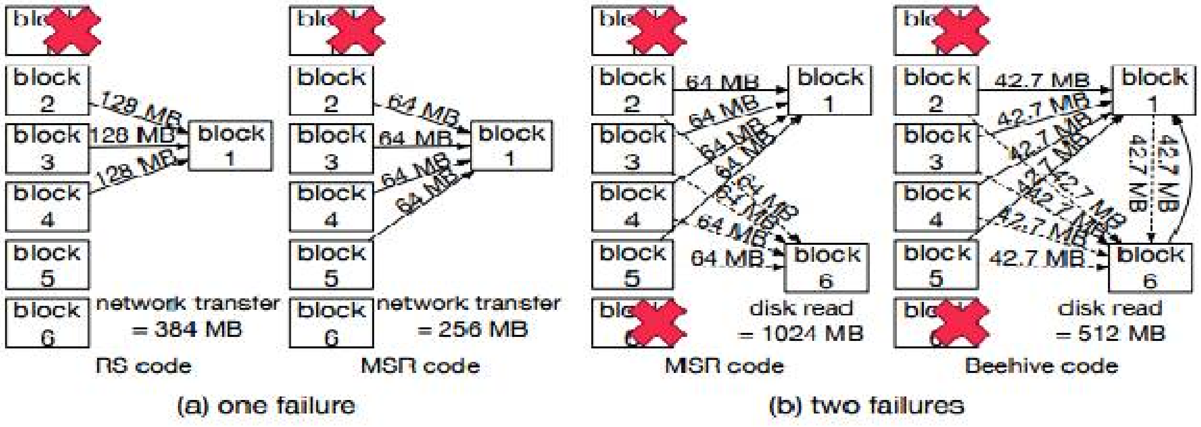
Setup phase consists of users having their files to store with unique file identifier knownis Im.Also if we use separate storage and key servers then it includes configuration ofthem.Encryption and erasure codes are added in this phase.

Storage phase consists of Storage servers at different location, their storage policies, installationof prerequisites like code or hardware.

Integrity Phase calculates integrity tag code with the help of verification key which is storedon key servers.

And In last phase data is reconstructed at all failed disks simultaneously using Beehive erasures.As name suggests, Beehive is like nest of bees and they reconstruct it cell by cell.All bees start 1 cell together hence construction is easy and less time consuming.

Retrievalphase consists of activities like retrieving files with decryption and checking their integritywith verification keys.



An instant benefit this brings is that each block will only be read once to reconstruct multipleblocks. As illustrated in above figure, the total amount of disk read is saved by 50% when we reconstruct two blocks togetherat the same time, while we can even further save network transfer as well. In fact, Beehive codes achieve the optimal network transfer per block in the reconstruction operation. The construction of Beehive codes is built on top of MSR codes.

- Mathematical Model

Let S be the system of beehive erasure codes.
S = (k, r, d,t,B,α, ω)
Where
k- Original data blocks
r - Parity blocks
d- No.of storage servers
t- No.of unavailable blocks ¿1
B- Total size of data
α = d-k+1 (no. of segments in k)
ω = B/ (α k) bytes of data in each segment.

B. *Observations:*

This system is having very high robustness and security compared to error correction codes and other encryptionalgorithms, Array codes or RAID levels.

IV. ADVANTAGES

1) The tight integration of encoding, encryption, andBeehive codes makes the storage system efficiently meet therequirements of data robustness and Integrity checks would add data confidentiality.
2) More flexible adjustment between the number of storageservers and robustness.
3) As Beehive codes can reconstruct data simultaneously at all failed servers it works with optimized network transfer. Hence latency is reduced and system is up and running within timeliness.

V. Applications

1. Government Organizations such as defence sector, spyorganizations where high data confidentiality and securityis required.
2. Space research centres
3. Nuclear power plants.
4. Our service can be used by the Data Centres on clouds forthe storage of the users data and also forward securely.
Our service can be used by the business organizationsfor maintaining the integrity and secrecy of the importantinformation.
5. Spy organizations those work for our nation need to be verysecure due to highly sensitive data.
6. Nuclear power plants are also need to be taken care becausea smallest mistake can harm many people.
7. In banking domain, we may provide integrity check schemefor KYC documents management.
8. Our system is highly distributed where storage servers independentlyencode and forward messages and key serversindependently perform partial decryption.

VI. Conclusion and Future Work

The implementations of the traditional systems may face crashes, DOS attacks and unavailability due toregional network outages. In the proposed system a secure distributed storage systemis formulated by integrating an encryption scheme with adecentralized beehive erasure code.
Proposed scheme supports not only the expected encodingoperations but also optimized network transfer with slightly bandwidth overhead.
Enhanced robustness of system is achieved by Integritychecks at low cost and compatible manner.
Future scope:
1) To use efficient algorithms for encryption and decryptionpurpose
2) To implement storage servers and key servers to thecloud for better modularity.
3) To generate integrity tags whileencryption.

VII. Acknowledgment

I take this opportunity to express my heartfelt gratitudeto my guide and head of department, Prof.Vina M Lomte, Department of Computer Engineering, RMDSSOE, SavitribaiPhule Pune University, for her kind cooperation, constant Encouragement and suggestions and capable guidance duringthe research, without which it would have been difficult to proceed with.

References

1. Jun Li and Baochun Li,"Beehive:Erasure Codes for Fixing Multiple Failures in Distributed Storage Systems", IEEE Transactions on Parallel and Distributed Systems,Vol. 64, no.3 , pp. 840 - 851, April-March. 2017
2. Shiuan-Tzuo Shen; Hsiao-Ying Lin; Wen-Guey Tzeng,"An Effective Integrity Check Scheme for Secure Erasure Code-Based Storage Systems", IEEE Transactions on Reliability,Vol. 64, no.3 , pp. 840 - 851,June.2016
3. Laszlo Czap,Christina Fragouli,Vinod Prabhakaran, Suhas Diggavi,"Secure Network Coding With Erasures and Feedback", IEEE Transaction on InformationTheory ,April.2015
4. Jian Liu,Kun Huang,Hong Rong,Huimei Wang and Ming Xian,"Efficient and Secure Data Forwarding for Erasure-code based Cloud Storage", Proc 2015 IEEE Intl Workshop on cloud computing system,Networks and applications
5. H.-Y. Lin and W.-G. Tzeng,"A secure erasure code-based cloud storage system with secure data forwarding", IEEE Trans. Parallel Distrib. Syst.,Vol. 23, no. 6, pp. 9951003, Jun. 2012
6. H.-Y. Lin and W.-G. Tzeng,"A secure decentralized erasure code for distributed networked storage", IEEE Trans. Parallel Distrib. Syst., Vol. 21, no. 11, pp. 15861594, Nov. 2010.
7. C. Wang, Q. Wang, K. Ren, and W. Lou,"Ensuring data storage security in cloud computing", in Proc. 17th Int. Workshop Quality of Service (IWQoS'09), Jul. 2009, pp. 19.
8. G. Dimakis, V. Prabhakaran, and K. Ramchandran,"Decentralized erasure codesfor distributed networked storage," IEEE Trans. Inf.Theory, Vol. 52, no. 6, pp. 28092816, Jun. 2006.

BIOGRAPHY

Mugdha Adivarekaris a Student in the Computer Engineering Department, RMD Sinhgad School of Engineering,Warje, PuneUniversity. She is pursuing Master of Computer engineering degree in. Her research interests are Information Retrieval, Data mining, Algorithms, etc.