ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

SECURING DATA IN THE BIG DATA STORAGE USING ERASURE CODED

S.Krishnakumar

Research scholar, Department of Computer Science, Govt. Arts College, Ariyalur, Tamilnadu, India.

P.Selvakumar

Research Supervisor, Asst. Prof. of Computer Science, Govt. Arts College, Ariyalur, Tamilnadu, India.

ABSTRACT

Modern distributed storage systems present immense capacity to gratify the exponentially growing need of storage space. They frequently use erasure codes to defend against disk and node failures to supplement reliability, while difficult to accumulate the latency requirements of the applications and clients. This paper provides a perceptive higher bound on the standard service delay of such erasure-coded storage with random service time distribution and consisting of multiple heterogeneous files. Not only does the product be successful known delay bounds that only vocation for a particular file or homogeneous files, it also enables a unique problem of joint latency and storage cost minimization over three dimensions: selecting the removal code, placement of determined chunks, and optimizing scheduling policy. The difficulty is competently solved via the calculation of a sequence of convex approximations with provable convergence. We additional model our solution in an open-source cloud storage deployment over three geographically distributed data centers. Experimental results authenticate our theoretical delay investigation and demonstrate significant latency reduction, providing valuable insights into the projected latency–cost tradeoff in erasure-coded storage.

KEYWORDS: content placement, data center, difference- Of-convex programming, distributed storage, erasure code, Gradient descent, joint optimization, latency

INTRODUCTION

A quantification of service latency for erasure-coded storage with frequent heterogeneous files and proposes an imaginative solution to the joint optimization of both latency and storage cost. The result of coding on contented retrieval latency in data-center storage systems is depiction more and more significant consideration these days, as Google and Amazon have published that every 500 ms extra delay means a 1.2% user loss. However, to our best information quantifying the exact service delay in an erasure-coded storage system is an open problem, prior works focusing on asymptotic queuing delay behaviors are not applicable because redundancy factor in practical data centers classically stay small due to storage cost concerns. Due to the lack of analytical latency models for erasure-coded storage, most of the literature is focused on dependable distributed storage system design, and latency is only obtainable as a performance metric when evaluating the proposed removal coding scheme, e.g, which show latency development due to removal coding in different system implementations. Related design can also be found in data access scheduling access collision avoidance, and encoding/decoding time optimization and there are also some works using the LT erasure codes to adjust the system to get together user requirements such as availability, integrity, and confidentiality. Restricting to the particular case of a single file or homogeneous files, service delay bounds of erasure-coded storage have been newly studied.

Queuing-Theoretic Analysis: For a single file or multiple but homogeneous files, under an supposition of exponential service time distribution, the author in proved an asymptotic result for symmetric large-scale systems that can be functional to give a assessable approximation for expected latency, however, under a

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

statement that chunk placement is fixed and so is coding policy for all requests, which is not the case in reality. Also, the authors in and planned a block-one-scheduling policy that only allows the demand at the head of the buffer to move forward. An upper bound on the average latency of the storage system is provided through queuing-theoretic analysis for MDS codes with. Later, the approach is extended in to general erasure codes, yet for a single file or homogeneous files. Families of MDS-Reservation scheduling policies that block all except the first of file requests are proposed and lead to arithmetical upper bounds on the average latency. It is shown that as increases, the bound becomes tighter while the number of states concerned in the queuing-theoretic analysis grows exponentially

Consumers are engaged in further social networking and E-commerce activities these days and are ever more storing their documents and media in the online storage. Businesses are relying on Big Data analytics for business intelligence and are migrating their customary IT infrastructure to the cloud. These trends cause the online data storage demand to rise faster than Moore's Law. The augmented storage demands have led companies to begin cloud storage services like Amazon's S3 and personal cloud storage services like Amazon's Cloud drive, Apple's iCloud, Drop Box, Google Drive, Microsoft's Sky Drive, and AT&T Locker. Storing redundant information on distributed servers can amplify reliability for storage systems since users can repossess duplicated pieces in case of disk, node, or site failures. Erasure coding has been broadly studied for disseminated storage systems and used by companies like Facebook and Google since it supply space-optimal data redundancy to secure beside data loss

RELATED WOEKS

In [1] Guanfeng Liang, and Ulas¸ C. Kozat et al presents our paper presents solutions that can significantly get better the delay performance of putting and retrieving information in and out of cloud storage. We first center on measuring the delay performance of a very accepted cloud storage service Amazon S3. We create that there is important randomness in service times for appraisal and characters small and medium size objects when assigned distinct keys. We further demonstrate that using erasure coding, parallel connections to storage cloud and limited chunking (i.e., dividing the object into a few smaller objects) together pushes the envelope on service time distributions significantly. Thus, in the second fraction of our paper we focus on analyze the delay performance when chunking, FEC, and parallel connections are used together. Based on this analysis, we develop load adaptive algorithms that can pick the best code rate on a per demand basis by using off-line computed queue backlog thresholds.

In [2] systems akshay kumar, ravi tandon, t. Charles Clancy et al presents Distributed (Cloud) Storage Systems (DSS) show heterogeneity in several dimensions such as the vole (Size) of data, frequency of data access and the preferred amount of reliability. Ultimately, the multifaceted between these dimensions impacts the latency performance of cloud storage systems. To this end, we suggest and examine a heterogeneous distributed storage model in which n storage servers (disks) hoard the data of R distinct classes. Data of class i is encoded using a (n; ki) erasure code and the (random) data retrieval needs can also vary from class to class. We present a queuing theoretic analysis of the proposed model and create upper and inferior bounds on the average latency for each data class under a range of preparation policies for data retrieval.

In [3] B. Rex Cyril, DR. S. Britto Ramesh Kumar et al presents Data security has a major issue in cloud computing environment; it becomes a severe problem due to the data which is stored diversely over the cloud. Data privacy and security are the two main aspects of user's anxiety in cloud information technology. Numerous techniques regarding these aspects are gaining consideration over the cloud computing environments and are examine in both industries and academics. Data privacy and security protection are becoming the most important aspect for the future enhancement and development of cloud computing technology in the field of business and government

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

sectors. Thus, in this paper, the cloud computing safety techniques are assessed and their challenges regarding data protection are discussed.

In [4] Alexandros G. Dimakis, Kannan Ramchandran, Yunnan Wu et al presents Distributed storage systems frequently commence redundancy to supplement reliability. When coding is used, the reinstate problem arises: if a node storing prearranged in sequence fails, in order to continue the comparable level of reliability we need to produce encoded information at a new node. This amounts to a partial recuperation of the code, whereas conventional removal coding focuses on the complete recovery of the information from a subset of encoded packets. The consideration of the refurbish network traffic gives rise to new design challenges. Recently, network coding techniques have been involved in address these challenge, establishing that preservation bandwidth can be summary by orders of magnitude compared to normal erasure codes

In [5] Virag Shah, Gustavo de Veciana et al presents Large scale Content Delivery Networks (CDNs) is one wherein servers can work mutually, as a pooled resource, to meet personality user requests. In such systems basic questions include: How and where to replicate files? What is the authority of dynamic service share across request types, and whether it can supply extensive gains over simpler load balancing policies? What are tradeoffs between performance, reliability and recovery costs, and energy? The document provides both plain and asymptotic approximations for great systems towards addressing these essential questions of the enter components of today's information infrastructure

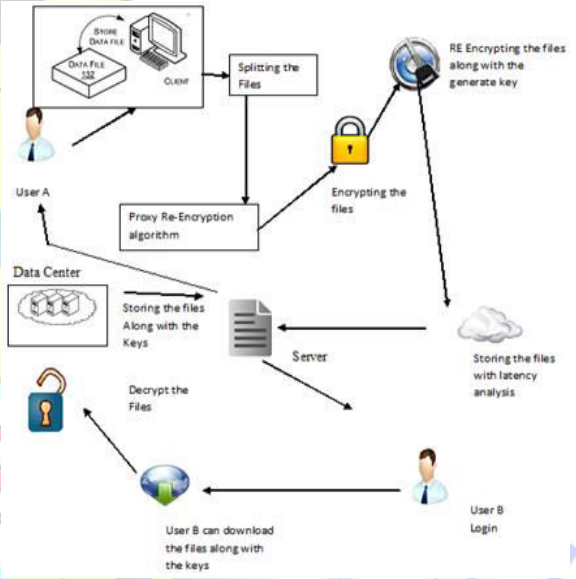
PROPOSEDSYSTEM

In the propose a methodical framework that: 1) quantifies the outer bound on the service latency of arbitrary removal codes and for any numeral of files in distributed data center storage with universal service time distributions; and 2) enables a novel answer to a joint minimization of latency and storage cost by optimizing the system over three dimensions: erasure coding, chunk placement, and scheduling policy.

MODULE SPECIFICATION

- Decentralized erasure code
- Proxy Re-Encryption
- Data storage phase
- Data forwarding phase
- Data retrieval phase

ARCHITECTURE DIAGRAM



DECENTRALIZED ERASURE CODE

In the decentralized erasure code is a removal code that independently computes each codeword sign for a message. Thus, the encoding procedure for a message can be whole into n parallel tasks of generating codeword symbols. A decentralized erasure code is suitable for employ in a distributed storage system. After the message symbols are sent to storage servers, each storage server separately computes a codeword symbol for the received message symbols and stores it. This finishes the encoding and storing progression. The recovery development is the same.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

PROXY RE-ENCRYPTION

In a proxy re-encryption scheme, a proxy server can relocate a cipher text beneath a public key PKA to a novel one under an additional public key PK_B by using the re-encryption key $RK_{A->B}$. The server does not recognize the plaintext during transformation. In proposed some proxy re-encryption schemes and practical them to the allocation function of protected storage systems. In their work, messages are first encrypted by the possessor and then stored in a storage server. When a user needs to split his messages, he sends a re-encryption key to the storage server. The storage server re-encrypts the encrypted messages for the authorized user. Thus, their system has data confidentiality and supports the data forwarding purpose. Our work additional integrates encryption, re-encryption, and encoding such that storage robustness is strengthened.

DATA STORAGE PHASE

When user A needs to accumulate a message of k blocks m1,m2,. . .;mk with the identifier ID, he computes the individuality token and performs the encryption algorithm Enc k blocks to get k original cipher texts C1,C2, . . . ; Ck. A unique cipher text is indicated by a leading bit b ¼ 0. User A sends each cipher text Ci to v erratically chosen storage servers. A storage server receives a position of inventive cipher texts with the equivalent identity token _ from A. When a cipher text Ci is not conservative, the storeroom server inserts Ci to the set. The person format of is a mark for the absence of Ci. The storage server performs Encode on the set of k cipher texts and stores the encoded result (codeword symbol). Encryption. Encoding is major division in the data storage

DATA FORWARDING PHASE

User A wants to onward a message to one more user B. He wants the first component a1 of his secret key. If A does not possess a1, he queries key servers for key shares. When at least t key servers respond, A recovers the first module a1 of the secret key SK_A via the Key Recover algorithm. Let the identifier of the message be ID. User A computes the re-encryption key $RK^{ID}_{A->B}$ via the Re Key Gen algorithm and strongly sends the re encryption key to each storage server. By using $RK^{ID}_{A->B}$ a storage server re-encrypts the unique codeword symbolC0with the identifier ID into a re-encrypted codeword symbolC'' via the ReEnc Þ algorithm such that C'' is decrypt able by using B's secret key. A re-encrypted codeword symbol is indicated by the most important bit b ¼ 1. Let the public key PK_B of user B be $(g^{b1}; h^{b2}.)$

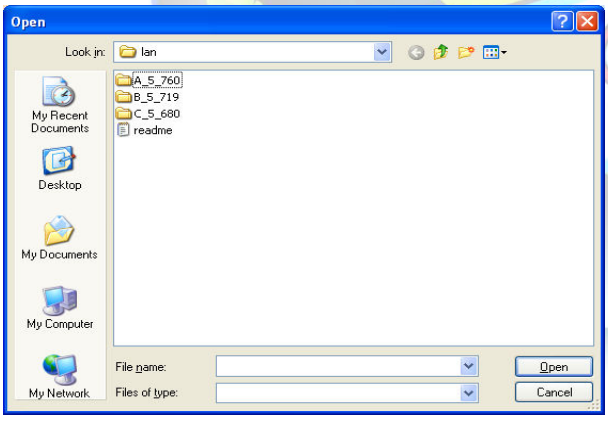
DATA RETRIEVAL PHASE

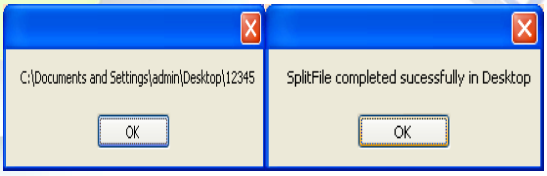
There are two belongings for the data retrieval phase. The first case is that a user A retrieves his possess message. When user A needs to recover the message with the identifier ID, he informs all type servers with the identity token A key server first retrieves inventive codeword symbols from u randomly chosen storage servers and then performs limited decryption Share Dec on each retrieved original codeword symbolC0. The result of partial decryption is called a partially decrypted codeword symbol. The key server sends the incompletely decrypted codeword symbols _ and the coefficients to user A. After user A collects replies from at least t key servers and at least k of them are initially from separate storage servers, he execute Combine on the t partially decrypted codeword symbols to recover the blocksm1;m2; . . .;mk. The next case is that a user B retrieves a message forwarded to him. User B informs all key servers straight. The collection and combining parts are the similar as the first case excluding that key servers retrieve reencrypted codeword symbols and achieve partial decryption Share-Decrypted on re-encrypted codeword symbols.

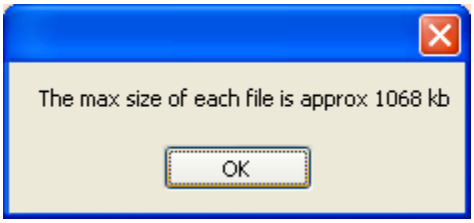
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

OUTPUT RESULTS

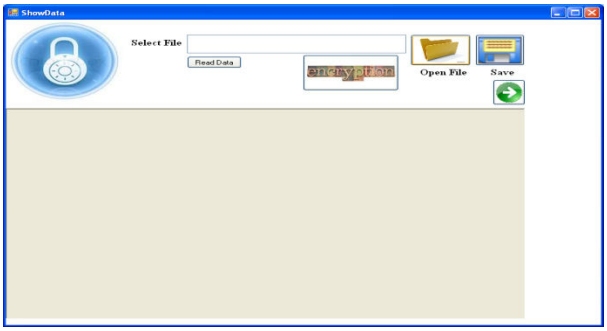


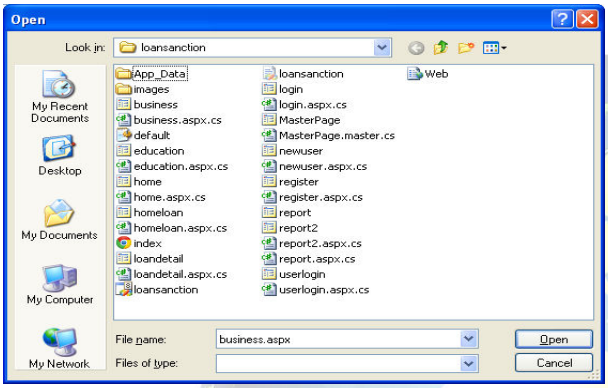


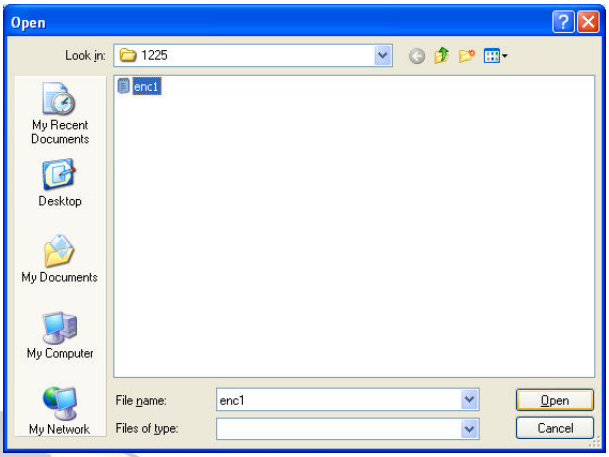


ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018







CONCLUSION

Relying on a narrative probabilistic scheduling policy, this paper develops an investigative higher bound on average service delay of erasure-coded storage with arbitrary amount of files and any service time distribution. A joint latency and cost minimization is formulated by cooperatively optimizing over erasure code, chunk placement, and preparation policy. The minimization is solving using an efficient algorithm with proven convergence. Even though only local optimality can be certain due to the nonconvex nature of the mixed-integer optimization problem, the proposed algorithm considerably reduces a latency-plus-cost purpose. Both our conjectural analysis and algorithm design are validated via an example in Tahoe, an open-source dispersed file system.

REFERENCE

[1] E. Schurman and J. Brutlag, "The user and business impact of server delays, additional bytes and http chunking in web search," presented at the OReilly Velocity Web Perform. Oper. Conf., Jun. 2009.

[2] G. Liang and U. Kozat, "FAST CLOUD: Pushing the envelope on delay performance of cloud storage with coding," IEEE/ACM Trans. Netw., vol. 22, no. 6, pp. 2012–2025, Nov. 2013.

[3] S. Chen et al., "When queueing meets coding: Optimal-latency data retrieving scheme in storage

clouds," in Proc. IEEE INFOCOM, Apr. 2014, pp. 1042–1050.

[4] G. Liang and U. C. Kozat, "TOFEC: Achieving optimal throughputdelay trade-off of cloud storage using erasure codes," in Proc. IEEE INFOCOM, Apr. 2014, pp. 826–834.

[5] V. Shah and G. Veciana, "Performance evaluation and asymptotic for content delivery networks," in Proc. IEEE INFOCOM, Apr. 2014, pp. 2607–2615.

[6] C. Angllano, R. Gaeta, and M. Grangetto, "Exploiting rate less codes in cloud storage systems," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1313–1322, May 2015.

[7] A. Kumar, R. Tandon, and T. C. Clancy, "On the latency of erasure coded cloud storage systems," arXiv:1405.2833, May 2014.

[8] A. D. Luca and M. Bhide, Storage Virtualization for Dummies, Hitachi Data Systems Edition. Hoboken, NJ, USA: Wiley, 2009.

[9] Amazon S3, "Amazon Simple Storage Service," [Online]. Available: http://aws.amazon.com/s3/

[10] M. Sathiamoorthy et al., "XORing elephants: Novel erasure codes for big data," in Proc. 39th VLDB Endowment, 2013, pp. 325–336.