Optimal Security & Performance In File Hosting Applications

Sukruti Admuthe[1], Prof. Bharti Dhote[2]

Dept. Computer Computer Engineering,
SIT , Lonavala Pune, India,
sukruti.admuthe@gmail.com[1] , bld.sit@sinhgad.edu[2]

---***---

Abstract - *Individual distributed storage administrations are picking up fame. However the two principle issues worried with the document facilitating applications are security and execution. Through this anticipate for information security "Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) system" would be utilized. In the DROPS strategy, the document is partitioned into sections, and repeat the divided information over the cloud hubs. Each of the hubs will stores just a solitary section of a specific information document that guarantees that even if there should be an occurrence of an effective assault, no important data is uncovered to the aggressor .Also considering the purpose of replication, Erasure coding offers better information assurance. Exploratory results exhibit that the system is better to the extent relieved storage room limit and data recuperation.*

Key Words: EC2, S3, DROPS, Centrality, T-Coloring , Erasure Coding.

I . INTRODUCTION

Distributed computing pulls in an unfathomable enthusiasm from both industry and the scholarly world, serving as engineering stage for an assortment of administrations. Distributed storage administrations, specifically, are picking up fame among both residential and undertaking clients as a basic, handy and safe instrument to store information. Such fame keeps on expanding with the late passage of huge players, for example, Google and Microsoft, to the distributed storage market.

The cloud computing paradigm has reformed the usage and management of the information technology infrastructure. Cloud computing is characterized by on-demand self-services, ubiquitous network accesses, resource pooling, elasticity, and measured services. The aforementioned characteristics of cloud computing make it a striking candidate for businesses, organizations, and individual users for adoption. However, the benefits of low-cost, negligible management (from a user's perspective), and greater flexibility come with increased security concerns. The cloud computing worldview has changed the utilization and management of the information technology framework. Cloud computing is described by on-demand self-services, universal ubiquitous network accesses, resource pooling, elasticity, and measured services. The same qualities of cloud computing make it a striking possibility for organizations, associations, and individual clients for appropriation. Be that as it may, the upsides of economical, negligible management (from a client's point of view), and larger flexibility escort enhanced security considerations.

Security is a standout amongst the most vital viewpoints among those disallowing the across the board selection of cloud computing. Cloud security issues might stem because of the core technology implementation (virtual machine (VM) escape, session riding, and so on.), cloud service offerings (structured query language injection, weak authentication schemes, etc.), and emerging from cloud qualities (data recovery vulnerability, Internet protocol vulnerability, etc.). For a cloud to be safe for data, the greater part of the taking interest elements must be secured. In any given framework with various units, the most abnormal amount of the framework security is equivalent to the security level of the weakest entity. In this manner, in a cloud, the security of the elements does not exclusively rely on upon an individual's security measures. The neighboring elements might give a chance to an attacker to sidestep the client's defenses.

The rest of this paper is organized as follows: In Section II, we present the related works. Section III the key terminologies .After that, we examine the existing system model and design the proposed system model Section IV. Section V presents the algorithms implemented in project with modifications. Practical mathematical issues are discussed in Section VI. Section VII further evaluates its performance and Section VIII concludes the paper.

II . RELATED WORK

In paper [1] Division and Replication of Data in the Cloud method is elaborated for Optimal Performance and Security (DROPS) that collectively approaches the security and performance issues. Outsourcing information to an outsider authoritative control, as is done in distributed computing, offers ascend to security concerns. The information bargain may happen because of assaults by different clients and hubs inside the cloud. In the DROPS methodology, we divide a file into fragments, and replicate the fragmented data over the cloud nodes. Each of the nodes stores only a single fragment of a particular data file that ensures that even in case of a successful attack, no meaningful information is revealed to the attacker. Moreover, the nodes storing the fragments, are separated with certain distance by means of graph T-coloring to prohibit an attacker of guessing the locations of the fragments.

In paper [2] present an underlying estimation to comprehend the outline and execution bottleneck of the Dropbox framework. Our estimation distinguishes the cloud servers/occurrences used by Dropbox, uncovering that it not just depends on Amazon's S3 for document stockpiling, additionally uses Amazon's EC2 examples to give such key capacities as synchronization and coordinated effort. This half and half plan makes viable utilization of cloud assets for both calculation (with EC2) and capacity (with S3), accordingly empowering consistent coordinated effort and record synchronization among various clients .

In paper [3],to solve the problem of network scalability and oversubscription authors worked on comparison of the major DCN architectures. They also simulated performance of the important DCN architectures in numbers of realistic conditions that to in various network configurations. The outcomes prove that in terms of average network throughput and packet delay the fat-tree based DCN architecture is not performed up to the mark the DCell and three-tier DCN architectures.

In paper [4] authors examines the structural robustness of the state-of-the-art data center network (DCN) system. After examination the outcomes showed that the DCell architecture debases smoothly under the greater part of the failure types when contrasted with the FatTree and ThreeTier design and also the outcome from deterioration metric shows the DCell is the robust architecture among all of the considered DCNs.

In paper [5], authors explained a new method which secure could data by ensuring range of protection from integrity and freshness verification to high data availability. Authors also present an auditing system which gives tenants visibility into proper operation of cloud.

III . Key Terminologies

The Before we delve into the points of interest of the DROPS technique, we present the related ideas in the accompanying for the simplicity of the users.

- EC2 - Amazon Elastic Compute Cloud (EC2) shapes a focal piece of Amazon.com's distributed computing stage, Amazon Web Services (AWS), by permitting clients to lease virtual PCs on which to run their own PC applications.
- S3 – The Amazon Simple Storage Service (Amazon S3) is an adaptable, fast, minimal effort Web-based administration intended for online reinforcement and chronicling of information and application programs.
- DROPS - The Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that all things considered methodologies the security and execution issues. In the DROPS procedure, the document is partitioned into sections, and duplicate the divided information over the cloud hubs.
- Erasure Coding - EC is a technique for information security in which information is broken into sections, extended and encoded with excess information pieces and put away over different locations or storage media.
- DES Algorithm - The Data Encryption Standard (DES) is a symmetric key block cipher and an implementation of a Feistel Cipher for encryption of the data present in the file.
- Closeness Centrality - A node is said to be closer with respect to all other nodes within a network, if sum of the distance from the other entire node is lower than sum of the distance of other candidate nodes from all of the other nodes.
- T-coloring - Suppose we have a graph G(V,E)and a set of T containing non negative integer including 0, T-coloring is the mapping of function from a vertex of V to a set of non negative integer such a that | f(x) - f(y) | T where (x , y)E. The mapping function f assigns color to a vertex and the distance between color to the adjacent vertices must not belong to T.

III . Implementation Details

In this paper we propose not to store the entire file at a single node. The DROPS methodology fragments the file and makes use of the cloud for replication. We display Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that judicially parts client documents into pieces and duplicates them at key areas inside the cloud. The division of a record into parts is performed in view of a given client criteria such that the individual sections don't contain any important data. Each of the cloud hubs (we utilize the term hub to speak to processing, stockpiling, physical, and virtual machines) contains an unmistakable section to build the information security.

A . System Architecture

Our main goal is to propose Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) and Erasure Coding that collectively approaches the security and performance issues. The fragments are distributed such that no node in a cloud holds more than a single fragment, so that even a successful attack on the node leaks no significant information. The Erasure methodology uses controlled replication where each of the fragments is replicated only once in the cloud to improve the security.In the DROPS methodology, user sends the data file to cloud.

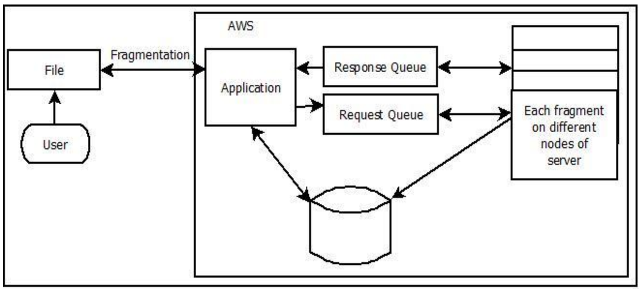


Fig -1: Proposed Architecture

The cloud manager system (a user facing server in the cloud that entertains user's requests) upon receiving the file performs: (a) fragmentation, (b) first cycle of nodes selection and stores one fragment over each of the selected node, and (c) second cycle of nodes selection for fragments replication. The fragment of file is done by Erasure encoding in such a way that a file is first divided into four parts and next two part will contain the redundant data piece , which are then stored across six different locations with centrality measures i.e. those fragment are stored to nearest six servers among available servers. The cloud manager keeps record of the fragment placement and is assumed to be a secure entity.

B. Algorithms

The DES satisfies both the desired properties of block cipher. These two properties make cipher very strong.

- Avalanche effect – A small change in plaintext results in the huge grate change in the ciphertext.
- Completeness – Each bit of ciphertext relies on many bits of plaintext.

DES is an implementation of a Feistel Cipher. It uses 16 round Feistel structure. The block size is 64-bit. Though, key length is 64-bit, DES has an effective key length of 56 bits, since 8 of the 64 bits of the key are not used by the encryption algorithm (function as check bits only).

Algorithm 1 : Algorithm for security of data (DES)

1. P-boxes - The initial and final permutations are straight Permutation boxes that are inverses of each other.
2. The DES function, f applies a 48-bit key to the rightmost 32 bits to produce a 32-bit output.
3. Then Expansion Permutation Box is implemented. Since right input is 32-bit and round key is a 48-bit, we first need to expand right input to 48 bits.
4. Whitener (XOR) – After the expansion permutation, DES does XOR operation on the expanded right section and the round key. The round key is used only in this operation.
5. Substitution Boxes. – The S-boxes carry out the real mixing (confusion). DES uses 8 S-boxes, each with a 6-bit input and a 4-bit output.
6. Straight Permutation – The 32 bit output of S-boxes is then subjected to the straight permutation with rule

We apply the Erasure Encoding technique here once the file is encrypted by the above DES Algorithm. By this Erasure encoding technique , we divide the encypted file into total six fragment where four of the fragment contains the actual data and remaining two fragment contains the redundant data.

Algorithm 2: Algorithm for fragmentation and replication of data.

1. Create six files in the same directory, breaking the input file into four and plus two parity shards.
2. Create a buffer holding the file size, followed by the contents of the file.
3. The data stored is the file size (four byte int), followed by the contents of the file, and then padded to a multiple of four bytes with zeros for all four data shards must be the same size.
4. Connect to the servers that are available and ready.
5. On basis of centrality select nearest six server locations to store the fragments and place the fragment to those locations.
6. On the basis of selected servers allocate the determined space from the buffer into the respective fragment and allocate to the selected server locations.

Algorithm 3: Algorithm for finding centrality location for fragment placement and replication of data.

Inputs and initializations:

```
O = {O1,O2 , ..... ON}
o = {sizeof(O1) , sizeof(O2) , ..... sizeof(ON)}
col = {open color , close color}
cen = {cen1 , cen2 , ......,cenM}
col ← open color¦ i
cen ←ceni for all  i
```

Compute:

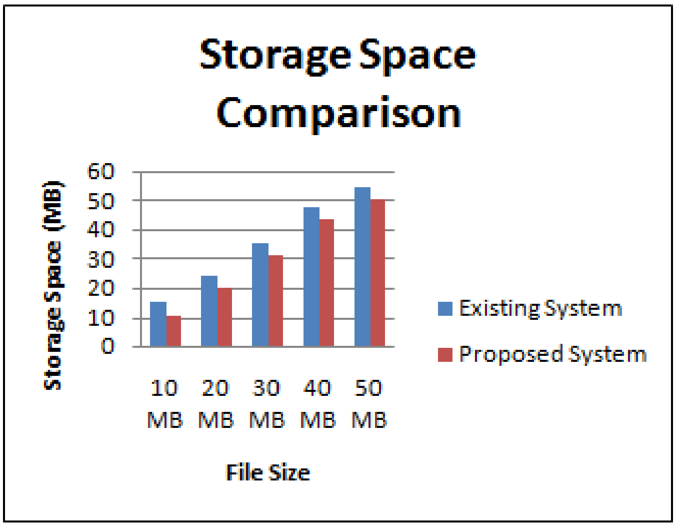
```
for each Ok in O do
    select Si | Si ← indexof(max(ceni))
    if colSi  = open color and si >= ok then
        Si     ← Ok
        si     ← si  - ok
        colSi  ←close color
        Si'    ←distance(Si; T)
        colSi  ←close color
    end if
end for
```

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

IV. Result

The project has been implemented in four sub modules. The first module provide security by DES Algorithm implementation by encrypting the file to be uploaded. The uploaded file is then fragmented by Erasure algorithm that provides in all four data fragments of file and two redundant data fragments in the second module , will helps to minimize the storage space by storing the redundant data fragment rather than all the four fragments .The next module consist of applying centrality theorem to find the nearest locations to save the fragments in minimum time.

This improvising the security and storage space for the good performance of the file hosting applications.



IV. Conclusion

To keep up the effective and to prevent information debasement in information storage backup system are main errands. Storing information sections on numerous servers decreases the possibilities of information loss yet this information section storage on different servers for information backup expands storage space. To preserve the storage space, our proposed framework executes erasure coding method; this can restore the tainted information records if any part is loss or corrupt. Likewise to decrease the calculation cost, framework makes utilization of cloud servers to store the information, as cloud server has its own particular focal points; security, low cost, high accessibility and so on. To assess the performance of proposed framework, the examination completed on dataset having numerous documents. The record size shifts from `16 MB to 11 MB. The exploratory results demonstrate that, our framework is superior to existing one, in case of storage space, cost and accessibility of information.

ACKNOWLEDGEMENT

We authors are thankful to the all researchers and publication houses as they have made their research information accessible for us. Teachers and PG co-coordinator for the cherished guidance. We are gratified to the members and faculty of Savitribai Phule University of Pune , for their persistent guidelines and helpful support. We are also thankful to the review examiner for their corrective ideas and suggestions. We are also thankful to the college faculty and authorities for providing the required infrastructure and helpful support. Lastly, we would like to prolong a sincere thanks to friends and family members.

REFERENCES

[1] Mazhar Ali , Kashif Bilal , Samee U. Khan , Bharadwaj Veeravalli , Keqin Li , Albert Y. Zomaya , “DROPS: Division and Replication of Data in Cloud for Optimal Performance and Security , ” IEEE Transactions on Cloud Computing , DOI 10.1109/TCC.2015.2400460

[2] Haiyang Wang, Ryan Shea, Feng Wang and Jiangchuan Liu , “On the Impact of Virtualization on Dropbox-like Cloud File Storage/Synchronization Services ,“ 978-1-4673-1298-1/12/$31.00!c 2012 IEEE

[3] K. Bilal, S. U. Khan, L. Zhang, H. Li, K. Hayat, S. A. Madani, N. Min-Allah, L. Wang, D. Chen, M. Iqbal, C. Z. Xu, and A. Y. Zomaya, “Quantitative comparisons of the state of the art data center architectures,” Concurrency and Computation: Practice and Experience, Vol. 25, No. 12, 2013, pp. 1771-1783.

[4] K. Bilal, M. Manzano, S. U. Khan, E. Calle, K. Li, and A. Zomaya, “On the characterization of the structural robustness of data center networks,” IEEE Transactions on Cloud Computing, Vol. 1, No. 1, 2013, pp. 64-77.

[5] A. Juels and A. Opera, “New approaches to security and availability for cloud data,” Communications of the ACM, Vol. 56, No. 2, 2013, pp. 64-73.

[6] G. Kappes, A. Hatzieleftheriou, and S. V. Anastasiadis, “Dike: Virtualization-aware Access Control for Multitenant Filesystems,” University of Ioannina, Greece, Technical Report No. DCS2013-1, 2013