International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 412 - 416, 2016, ISSN:- 2319–8656

Fragmentation of Data in Large-Scale System For Ideal Performance and Security

Dr. Shubhangi D.C
Department of Computer Science and Engineering, VTU Regional Centre Kalaburagi Karnataka, India.

Sonali V.Katke
Department of Computer Science and Engineering, VTU Regional Centre Kalaburagi Karnataka, India.

Abstract: Cloud computing is becoming prominent trend which offers the number of significant advantages. One of the ground laying advantage of the cloud computing is the pay-as-per-use, where according to the use of the services, the customer has to pay. At present, user's storage availability improves the data generation. There is requiring farming out such large amount of data. There is indefinite large number of Cloud Service Providers (CSP). The Cloud Service Providers is increasing trend for many number of organizations and as well as for the customers that decreases the burden of the maintenance and local data storage. In cloud computing transferring data to the third party administrator control will give rise to security concerns. Within the cloud, compromisation of data may occur due to attacks by the unauthorized users and nodes. So, in order to protect the data in cloud the higher security measures are required and also to provide security for the optimization of the data retrieval time. The proposed system will approach the issues of security and performance. Initially in the DROPS methodology, the division of the files into fragments is done and replication of those fragmented data over the cloud node is performed. Single fragment of particular file can be stored on each of the nodes which ensure that no meaningful information is shown to an attacker on a successful attack. The separation of the nodes is done by T-Coloring in order to prohibit an attacker to guess the fragment's location. The complete data security is ensured by DROPS methodology.

Keywords: Performance, Data Splitting, Cloud Security, T-Coloring, Data Security.

1. INTRODUCTION

The usage and management of the information technology infrastructure has reformed by the cloud computing model. Security plays the important aspect to prohibit the prevalent adoption of cloud computing. A cloud contains number of entities in it. In order to provide the security to the cloud, the participating entities have to be secure. Any organization considers the data to be the prime asset. The data must be secured when it is transferred to the public cloud, outside the organizations administrative domain. Accessing the data by the unauthorized users and process should be prevented. Otherwise the weak entity will put the cloud at risk. The prime concern has given to the data availability because of the data may move in the cloud which is not under the customer's administrative control. In cloud computing system, the data reliability problems, data availability problems, response time deal with the strategies of the replication. Data replicas are placed over the number of nodes which increases the attack of the particular data. Here we approach the issue of security and as well as performance as a problem of secure data replication. Within the cloud, the DROPS fragment the files and replicate them at the strategic locations. Based on the user criteria, the files are divided into fragments, and the individual fragment should not contain any meaningful information. In order to increase the data security in the cloud, each of the cloud nodes should contain distinct fragment. An attack on a single node will not show the location of the fragment in the cloud. In order to keep an attacker unsure about the file fragments location and to improve the security, the selection of the nodes is done in such a way that they must not be adjacent and must be at the certain distance from each other. The separation of the nodes is done by T-coloring. The selection of nodes is based on the centrality measures .To improve the retrieval time of the data which will ensure the improved access time. For the selection of the nodes, the two phases are performed. First, based on the centrality measures, the selection of the nodes is done for the initial placing of the fragments. Second, the replication is done for the selected nodes.

2. RELATED WORKS

The problem is the secure and optimal placement of data objects over distributive system within the network. Once the encryption key is divided into n shares and is distributed on different sites. The scheme (k,n) threshold secret sharing scheme is used to divide the key into n shares. The network can be divided into clusters. In every cluster, a primary site is selected which allocates the replicas in it. The scheme used here will combines the problem replication along with the security and improve the access time. This scheme focuses on providing the security to the encryption key. The fragmentation of the data files is not done and handles only single file [17].

The problem, to ensure the freshness, integrity and availability of the data within the cloud is by making use of the cryptographic techniques. The technique depends on the authentication scheme in order to provide the confidentiality of the data. Here the file blocks are stored in the various levels of the tree [6].

Information leakage is another problem that has been seen. In the case of improper sanitization and malevolent VM, the leakage of critical information is not handled. By exploiting the consolidated storage and native access control, the virtualized and multitenancy related issue regarding the cloud storage was approached here [7].

Everyone's perception of software delivery, infrastructure architectures and development models has been drastically altered by while ago emergence of cloud computing. The transition from main frame computer to the client or server implementation models, Cloud Computing includes the elements from grid computing, autonomic and utility computing into innovator implementation architecture. The quick transition towards the cloud has utilized the issue for the result of communication and information security. The Public Key Infrastructure (PKI) is utilized to improve the level of trust in integrity, authentication and confidentiality of data and communication between the involved parties. The certification authorities generate and manage the keys. For the storage of the keys temper proof devices are being utilized such as smart cards [18].

Likewise in [16] the public key cryptography is used and providing the data security in the cloud environment for the trusted third party. But in [16] the PKI infrastructure is not being utilized to minimize the overheads. The generation and management of private and public keys is the responsibility of the trusted third party which may be a single server or multiple servers. Protection can be provided to the symmetric keys by combining the public key cryptography and the (k,n) threshold secret sharing schemes. The dispositioning of the data and loss due to issues arising from virtualization and multi-tenancy cannot be protected under such schemes [16].

The security and optimal placement of data is done through the fragmentation and as well as the data objects replication. The fragmented file is encrypted and is stored within the network in a distributed fashion. In order for the increased availability of the data, replication is performed in a random manner [13].

Architectural and functional blocks of cloud computing which includes the data centers are immanent to the ICT sector (Information Communication Technology). The sundry hosts such as agriculture, nuclear science, and health care, search engines for research, smart grids, data storage and analysis. For cloud under building, a Data Center Networks (DCN) establishes the communicational backbone of a data center, discovering its performance boundaries. In order to overcome the failures, uncertainties in delivering the required Quality of Service level and to satisfy the service level agreement (SLA). The DCN should be strong. In view of manifold failure scenarios for performing a comparative analysis, includes the study of the classical robustness measures and also new procedures to quantitative the robustness of DCN. It also includes presenting multilayer graph modeling of various DCNs and also presents the inadequacy of classical network robustness measures for computing the robustness of the DCN [1].

The system that is an intrusion tolerant distributed system is designed in order that any forcible inclusion into the portion of the system will not endanger the availability, confidentiality and integrity. This approach is fit for distributed system therefore distribution empowers the separation of elements in order that forcible inclusion transfers physical access to only a portion of the system. By forcible inclusion, not only mean that, non-registered people can break-ins computer, but the registered users can also attempts to overstep or invective their immunity. Especially, possible administration security malice is taken into account. Few intensions of distributed systems can be designed to tolerate intrusions. In specific application function such as file management and security function such as user authorization and authentication [2].

The first and foremost aims to significant the major privacy, trust and the security issues in the current existing environment of cloud computing and also help users recognize the intangible and tangible threats associated with their users that includes the surveying for the most applicable security, trust and privacy issues that will pose the threats in the existing environment of cloud computing and also includes for analyzing the way which may be addressed to abolish these trust threats, security and potentiality privacy, and providing a believable, and dependable environment of cloud computing [15].

The proposed scheme that offloads the oftenly occurring dynamic authority generation operation on the trusted entity for keeping the minimum processing responsibility on mobile devices. Moreover, the proposed scheme compares with the existing one on the basis of time, performance metrics and energy consumption [11].

3. SYSTEM ARCHITECTURE

For a large-scale system's security, such as cloud depends on the security of the system and the security of and individual nodes. Thus the issue performance and security as a secure data replication problem is collectively approach. Here, within the cloud the fragment of the data files into pieces is done and replicates those fragments at strategic locations. Based on a user's given criteria, the file is divided into fragments, such that fragments should not include any meaningful information. Every node within the cloud holds a different fragment to raise the security of the data. After all there is the possibility of an effective attack on any node. To keep an attacker unsure about the fragments location and also to improve security, the selection of nodes should be in such a manner that they should not be adjacent and must be at the certain distance from each other. The separation of the nodes is ensured by using T-coloring. Based on the centrality measures, the nodes are selected and the retrieval time of the data is improved that will assure an improved access time. The fragments are replicated over the nodes to improve the retrieval time. The nodes are selected in two stages. In first stage, for initial placing of fragments, the nodes are selected based on the centrality measures. In second stage, for replication the nodes are selected. The data files fragmentation threshold is generated by the file owner. The fragmentation threshold can be specified by the file owner in terms of either percentage or number and size of the different fragments.

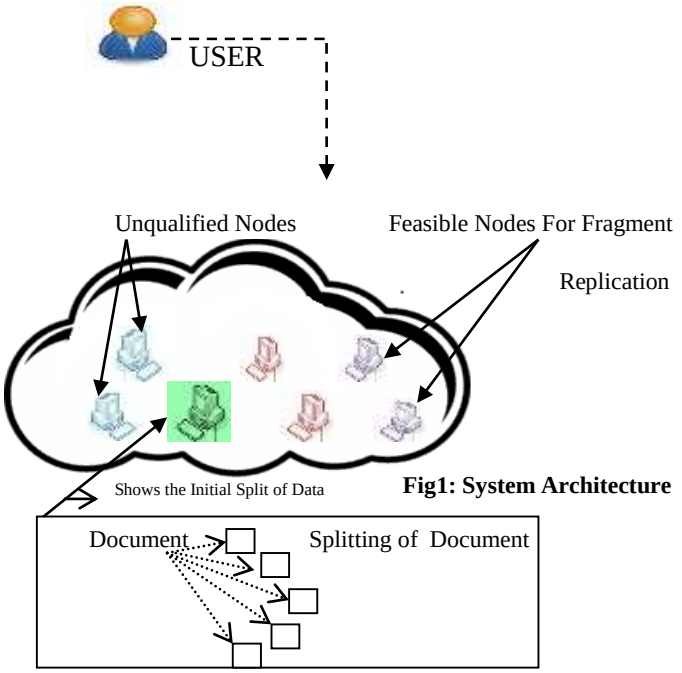


Fig1: System Architecture

For example, the percentage fragmentation threshold can dictate each fragment is of 5% size of the total file size. The owner may create a separate file that containing the information about the size and the fragment number for example, fragment one of size 5000 bytes, and the fragment two of size 8749 bytes. The file is best split by the owner such that each fragment should not contain important information as the owner is aware of the facts that pertaining to the data.

4. METHODOLOGY

To provide the security to the data file within the cloud the entire file is not stored on a single node. The DROPS methodology is used to fragment the file and replicate over the cloud. The fragments should be distributed such that only a node must store a single fragment, though the attack on the node will not be opened for an attacker. In order to improve the security within the cloud, the controlled separation is used by the DROPS methodology in which every fragment is replicated only once. In this methodology, the users need to send the data file to the cloud. Once receiving the data file, the cloud manager system will perform the fragmentation, then selects the node for storing the particular fragment and then replicates the fragments.

Fragment Placement Algorithm

$A = \{A1, A2 \ldots A_N\}$

a = {SIZE OF (A1), SIZE OF (A2)…...SIZE OF (A_N)}

COL = {OPEN_COLOR, CLOSE_COLOR}

CEN = {CEN1, CEN2,…….CEN $_M$}

COL←OPEN_COLOR FOR ALL i

CEN←CEN_i FOR ALL i

COMPUTE:-

FOR EACH A_k BELONGS TO A DO

SELECT P^i/P^i←INDEX OF (MAXIMUM (CEN_i))

IF COL s^i = OPEN_COLOR AND $pi >= a_k$ THEN

$P^i \leftarrow A_k$

$p_i \leftarrow p_i - a_k$

$COL_S{}^i$← CLOSE_COLOR

$P^{i'}$← DISTANCE (P^i,T)

$COL_P{}^i$ ← CLOSE_COLOR

END IF

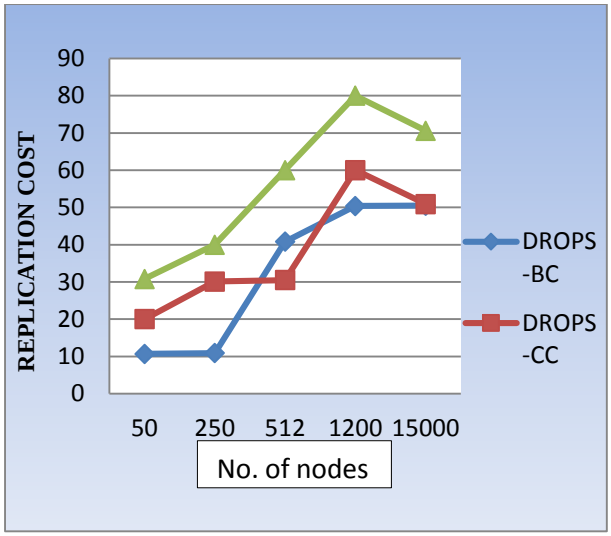
END FOR

5. RESULTS AND DISCUSSIONS

The DROPS methodology performance is compared with the algorithm. The algorithm's behavior is studied by incrementing the number of nodes in the system. By incrementing the number of nodes came to know about the performance of the placement techniques and the DROPS methodology. The Increment in the number of file fragments may strain the cloud storage capacity which in turn can alter the selection of the nodes.

Figure 2: Fragmentation of the data file to provide security.

The data file is divided into fragments, and these fragments are placed on the nodes, in order to provide the security to the data in the cloud.



igure3: Graph represents the centrality measures.

The graph shows the selection of the nodes with the maximum available storage capacity and the maximum centrality. It is obviously true by simple observation that the eccentricity centrality rise in the highest performance although the betweenness centrality resulted the lowest performance. The cause for this is that, in the network the nodes with the higher eccentricity are closer to all different nodes which results in the lower replication cost value for fragments access.

6. CONCLUSION AND FUTUREWORK

For many organizations transferring of data to the wireless servers has become a growing trend, because it takes away the heavy load of the maintenance and the local data storage. This work can consist the problem of generating the copies of data file and store those copies on the cloud servers. In the DROPS methodology, the scheme provided for the security of the cloud storage together as a whole deals with the performance and security in terms of the retrieval time. In order to provide the security to the data file in the cloud, the fragmentation of the data file was performed and the fragments were distributed over many nodes. The separation of nodes is done by using T-coloring. The dispersion and the fragmentation will assure that no important information will be procured by an opponent even on an effective attack. Within the cloud, the node should not store the multiple fragments of the file.

At present, with the methodology of the DROPS the file should be downloaded by the user, update the contents and again upload it. To update and identify the necessary fragments only an automatic update mechanism has to be developed. The time and the resources that are utilized in updating, downloading and again uploading the file will be saved by the above mentioned future work.

ACKNOWLEDGEMENT

We would like to take this opportunity to explicit our deep thankfulness to the Special Officer Dr. Baswaraj Gadge, Head of the Department Dr. Shubhangi D.C and the faculty members of computer science and engineering department of VTU PG CENTRE Kalaburagi for their inspiration and suggestions.

REFERENCES

[1] 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.
[2] Y. Deswarte, L. Blain, and J-C. Fabre, “Intrusion tolerance in distributed computing systems,” In Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy, Oakland CA, pp. 110-121, 1991.
[3] B. Grobauer, T.Walloschek, and E. Stocker, “Understanding cloud computing vulnerabilities,” IEEE Security and Privacy, Vol. 9, No. 2, 2011, pp. 50-57.
[4] K. Hashizume, D. G. Rosado, E. Fernndez-Medina, and E. B. Fernandez, “An analysis of security issues for cloud computing,”
Journal of Internet Services and Applications, Vol. 4, No. 1, 2013, pp. 1-13.
[5] W. A. Jansen, “Cloud hooks: Security and privacy issues in cloud computing,” In 44th Hawaii IEEE International Conference on System Sciences (HICSS), 2011, pp. 1-10.
[6] 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.
[7] 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.
[8] L. M. Kaufman, “Data security in the world of cloud computing,” IEEE Security and Privacy, Vol. 7, No. 4, 2009, pp. 61-64.
[9] S. U. Khan, and I. Ahmad, “Comparison and analysis of ten static heuristics-based Internet data replication techniques,” Journal of Parallel and Distributed Computing, Vol. 68, No. 2, 2008, pp. 113-136.
[10] A. N. Khan, M. L. M. Kiah, S. U. Khan, and S. A. Madani, “Towards Secure Mobile Cloud Computing: A Survey,” Future Generation Computer Systems, Vol. 29, No. 5, 2013, pp. 1278-1299.
[11] A. N. Khan, M.L. M. Kiah, S. A. Madani, and M. Ali, “Enhanced dynamic credential generation scheme for protection of user identity in mobile-cloud computing, The Journal of Supercomputing, Vol. 66, No. 3, 2013, pp. 1687-1706 .
[12] T. Loukopoulos and I. Ahmad, “Static and adaptive distributed data replication using genetic algorithms,” Journal of Parallel and Distributed Computing, Vol. 64, No. 11, 2004, pp. 1270-1285.
[13] A. Mei, L. V. Mancini, and S. Jajodia, “Secure dynamic fragment and replica allocation in large-scale distributed file systems,” IEEE Transactions on Parallel and Distributed Systems, Vol. 14, No. 9, 2003, pp. 885-896.
[14] L. Qiu, V. N. Padmanabhan, and G. M. Voelker, “On the placement of web server replicas,” In Proceedings of INFOCOM 2001, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 3, pp. 1587-1596, 2001.
[15] D. Sun, G. Chang, L. Sun, and X. Wang, “Surveying and analyzing security, privacy and trust issues in cloud computing environments,” Procedia Engineering, Vol. 15, 2011, pp. 2852 2856.
[16] Y. Tang, P. P. Lee, J. C. S. Lui, and R. Perlman, “Secure overlay cloud storage with access control and assured deletion,” IEEE Transactions on Dependable and Secure Computing, Vol. 9, No. 6, Nov. 2012, pp. 903-916.

[17] M. Tu, P. Li, Q. Ma, I-L. Yen, and F. B. Bastani, "On the optimal placement of secure data objects over Internet," In Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium, pp. 14-14, 2005.
[18] D. Zissis and D. Lekkas, "Addressing cloud computing security issues," Future Generation Computer Systems, Vol. 28, No. 3, 2012, pp. 583-592.
[19] J. J. Wylie, M. Bakkaloglu, V. Pandurangan, M. W. Bigrigg, S. Oguz, K. Tew, C. Williams, G. R. Ganger, and P. K. Khosla, "Selecting the right data distribution scheme for a survivable storage system," Carnegie Mellon University, Technical Report CMU-CS-01-120, May 2001.
[20] A. R. Khan, M. Othman, S. A. Madani, S. U. Khan, "A survey of mobile cloud computing application models," IEEE Communications Surveys and Tutorials, DOI: 10.1109/SURV.2013.062613.00160.