Secure Content De-Duplication Utilizing Efficient Content Discovery and Preserving De-Duplication (Ecdpd)

D.Vimala, S. Sangeetha, B. Sundar Raj

Abstract: Cloud computing, an efficient technology that utilizes huge amount of data file storage with security. However, the content owner does not controlling data access for unauthorized clients and does not control data storage and usage of data. Some previous approaches data access control to help data de-duplication concurrently for cloud storage system. Encrypted data for cloud storage is not effectively handled by current industrial de-duplication solutions. The deduplication is unguarded from brute-force attacks and fails in supporting control of data access .An efficient data confining technique that eliminates redundant data's multiple copies which is commonly used is Data-Deduplication. It reduces the space needed to store these data and thus bandwidth is saved. An efficient content discovery and preserving De-duplication (ECDPD) algorithm that detects client file range and block range of de-duplication in storing data files in the cloud storage system was proposed to overpower the above problems.Data access control is supported by ECDPD actively. Based on Experimental evaluations, proposed ECDPD method reduces 3.802 milliseconds of DUT (Data Uploading Time) and 3.318 milliseconds of DDT (Data Downloading Time) compared than existing approaches.

Keywords: Efficient content discovery and preserving De-duplication (ECDPD), De-duplication of data, Data Uploading Time(DUT), Data Downloading Time(DDT).

I. INTRODUCTION

De-duplication of data is a specialized data confiningmechanisms that eradicates the redundant data. For reducing data transmission rate in the network, the above method enhances both storage and network utilization . In the De-duplication process, depending on bytes specified by the client, transferred file is splitted into number of blocks in this process and then discoveresmatchless blocks of data file and is stored during the examination process. In the examination, other blocks were contrasted to the previously stored data copy. The matched block was substituted with a reference value to the saved block, whenever a match occurs.The redundant blocks may occur in this data file. Thus by using this technique, thestorage space and time can be reduced. Based on block size, match frequency can be measured.

Largestorage space is provided by various cloud services that maintains and control data file, which can contain files, texts, images etc. however, an industrial data de-duplication resolutions does not handle the encrypted data for cloud storage system. The previous data de-duplication resolutions are susceptible to brute-force assaults in storage system. A previous data de-duplication solution does not flexibly to help data access control and revocation for authorized clients. An existing data de-duplication does not provide security for data in cloud storage system. Duplicated file data can be stored by different cloud users at the server. As the storage space(cloud) is high, its utilized to waste networking assets, excess power is consumed, and makes data management difficult.

An efficient content discovery and preserving De-duplication (ECDPD) algorithm that detects client file range and block range of de-duplication in storing data files in the cloud storage system was proposed to overpower the above problems.Data access control is supported by ECDPD actively. The proposed system protects the secrecy of delicate data by sustainingdeduplication before outsourcing of data. Data security is protected by this system and attempts to formally address the problem of de-duplication of authorised data..The method prevents illegaldata files accessing and make duplicate file data on cloud server to store on cloud storage server after encryption of data file.The proposed system is identifying the unique data block which is stored in the cloud. Following are the contributions of this paper :

- Design an efficient content discovery and preserving De-duplication (ECDPD) algorithm which detecting client file range and block range of de-duplication in storing files in the cloud storage system
- To support authorized client's data access control security.
- To prevent illegal utilization of data files accessing and make duplicate file data on cloud storage server
- To reduce the interactive duplication discovery overheads and processing of data filesprocessin overheads
- To reduce the Uploading Time for data in a sec and Downloading Time for data in milliseconds contrast than previous methods.

II. RELATED WORKS

This A Cloud Computing Secure Framework (CCSF) which comprised four stages such as uniqueness management,

Revised Manuscript Received on October 22, 2019.

***** Correspondence Author

D.Vimala *, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.

S. Sangeetha, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.

B. Sundar Raj, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.

interruption discovery, and prevention method, data de-duplication, and secure data cloud storage was developed by Shobana et al. [1] .Kaaniche et al. [2] designed an OpenStack Swift that was a client-side de-duplication scheme for safely saving and allocation of externalised information through the public storage framework. Stanek et al. [3] evaluated a technique for encryption that guaranteed semantic protection stage for unpopular details offered lenient protection and enhanced storage ability and bandwidth benefits on relevant information. Akhila et al. [4] discussed and alteres a Data De-duplication approach as an easy data storage optimization mechanism in secondary then generally adopted in cloud storage region.

Harish et al. [5] developed a convergent encryption mechanism that utilized to overcome the data storage problems and to provide numerous protection mechanisms to particular data through verifying secret key. Thakar et al. [6] designed a hybrid cloud method that addressed a de-duplication occurring and supported authorized duplicate copy to validate in the hybrid cloud infrastructure. Shieh et al. [7] illustrated the concepts on de-duplication techniques and available de-duplication mechanisms such as pros and cons. It examined de-duplication mechanisms on the parameters like effectiveness, scalability, throughput, bandwidth ability and price.

Puzio et al. [8] developed a PerfectDedup for safe data de-duplication that considers data blockspopularity and levered the properties of hashing development that guarantees de-duplication. However, it fails to accomplish block level de-duplication. Priyadharsini et al. [9] examined various de-duplicationmetods to overcome the challenges. The de-duplication scheme diminished data storage demands in cloud computing environments with a significant VM floppy disk administrators quality. For a huge amount of VM disk supervisors, data storage client demands was failed to maintain.

Devi et al. [10] illustrated dissimilar methods that have been utilized data de-duplication in cloud storage framework and is generated by confining the data storage area needs for saving same type of data. Harnik et al. [11] discussed an easy approach that permitted cross-client de-duplication that reduces the thread of content discharge. It illustrates how de-duplication work in client network. Bharat et al. [12] discussed an approved content de-duplication and it protected information protection through the procedure of comprising disparity prerogative of the customers in the duplicate validation in the system.

III. PROPOSED SYSTEM

This The section represents methodology, steps for pre processing and the suggested system’s implementation. The suggested method is detecting client file range and block range of de-duplication in storing data files in the cloud storage system. It is utilized to eradicate duplicate photocopy of redundant data file. Figure 1 evinces the model of the suggested method with processing steps and mathematical assessment details. The pre-processing steps of implementation are shown in detail :

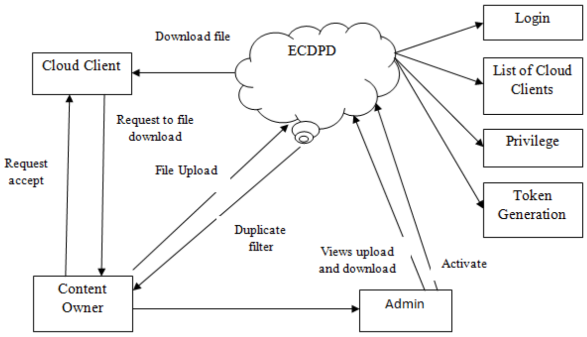

Fig 1 :Working Model of the Suggested ECDPD Algorithm

Authentication. In the module, the Content owner makes utilization of cloud assets to save, retrieve and send data file with various cloud clients. The content can be owned by an individual or an enterprise.

The uploading data file can be validated and neglated by .Content owner can view the de-duplicate file depend on cloud client can delete the redundant information. The data file can upload to the Cloud storage system from the content owner after that repeated file content upload is filtering de-duplication. An ECDPD algorithm applying the content owner side which utilizing filtering de-duplication.

Cloud Client. theuploading data file can be validated and neglated by .Content owner can view the de-duplicate file depend on cloud client can delete the redundant information. The data file can upload to the Cloud storage system from the content owner after that repeated file

The details can be registered by cloud client themselves and get the secret key for authentication and the cloud client can download the content owner’s uploaded data files.

The details can be registered by cloud client themselves and get the secret key for authentication and the cloud client can download the content owner’s uploaded data files. The cloud clients can ingress the file stored and depends on their access rights which are approvals granted by the content owner, like access rights store the data file in the cloud storage system.

Duplicate Checking. The cloud storage system is used to save content owner uploaded files. Data deduplication is a specialized data file compression mechanism that eliminates duplicate photocopy of replicating data file. Data compression and data storage system are inter-releated synonymous terms. The client files level and block level reduplications demanded by content owner before uploading data file. If no duplication is found, the data file is spitted into blocks and performs block level reduplication framework. Reduplication framework.

Data Distribution. Sharing and recovering is utilized in data distribution module. The data distribution is utilized for splitted and shared secret data. With sufficient data distribution is extracting and recovering the secret with the help of recovering method. Data distribution method partitions thedata file into similar size of blocks thatmakes equal size of random blocks and thentransmits into easy language.

Cloud Client Revocation. The administrator performs Cloud client revocation through accessible cloud client's revocation list and it based on which datafiles can be encrypted by content owner and assurance the confidentiality against the revoked cloud clients. The administratoralters the cloud client revocation recordeach day still no one has being revoked in the day.

Efficient content discovery and preserving De-duplication (ECDPD) algorithm. Efficient content discovery and preserving De-duplication (ECDPD) algorithm is upgrading client block level and file level de-duplication with dependability and distributing data documents with safely for cloud client's storage frameworks. A content owner gets a master key from eachoriginaldata. A duplicate and encodes the information duplicate with the master key. Furthermore, the client additionally infers a token for the information duplicate, such that the token will be utilized to identify duplicate copies. Here, we accept that the token accuracy property holds, i.e., if two information duplicates are the same, at that point their tokens are the same. To recognize duplicate copies, the client initially sends the token to the server side to verify if the indistinguishable duplicate has been already stored. Note that both the master key and the token are freely determined and the token can't be utilized to conclude the master key and negotiation information privacy. Both the encoded information duplicate copy and its comparing token will be stored on the server side. Formally, anECDPD plan can be characterized with four primitive capacities.

• KeyGen (D) → Data D to key K is mapped by K -key generation method.

• Enc(K,D) → Both data copy D and master key K were received from C -symmetric encryption method , then gives output cipher text C.

• Decce(K,C) → key K and cipher text C are inputs for decrypting method D, and the output of the D is provided.

• TokGen(D) → T(D) –maps M and gives output token T(D). ECDPD algorithm pseudo code is explained below in details:

Input: Any document file

Output: visualizedownloading time for data, and data uploading time

Procedure.

```
Start
Content owner authentication
Browse file to upload the cloud storage server
Apply ECDPD algorithm
Encrypt the original content file with token creation
Upload data file to cloud storage server process
If duplicate not present
Upload the data file to cloud storage server
Visualize data uploading time
Else
Compute Duplicate file and cannot upload the data file to cloud storage server
End if
Cloud client authentication process
Admin accept the cloud client authentication
Request content owner to download file
Content owner accept the cloud client request
Token send to requested cloud client
If token is correct
Cloud client downloads the requested file from cloud storage server
Else
Token is incorrect
Failed to download requested file
End if
End
```

Pseudo Code for Suggested Algorithm

IV. RESULTS & DISCUSSION

Experimental Setup. Intel Dual Core Processor with 1GB storage, and Window 7 system is used for deployment. The proposed ECDPD method is implemented in JAVA programming environment utilizing Netbeans 8.0, Apache Tomcat and MYSQL 5.5 database. The suggestedECDPD algorithm is evaluated with 2MB, 4MB and 8MB data.

Experimental Result. In this phase, suggested preserving and efficient information discovery. De-duplication (ECDPD) Algorithm represents a mathematical model as a well graphical view. The proposed ECDPD method is categorized in two parts where the first part elaborates mathematical equation of ECDPD methodology to design parameter for proposed approached evaluation. A second part represents the tabular and graphical result of ECDPD method according to various existing algorithms like a Leakage-Resilient (LR) [13], RandomizedSecure De-Duplication Scheme (SDS) [13] and Convergent Encryption (RCE) [13] . These all methods are tested with different parameters like data downloading time, and data uploading time. In the method is estimated with every parameter with various kinds of data separately.

Data Uploading Time. In the section, the approachsuggested enhances mathematical model for uploading time for data in equation (1). In the step, ECDPD method evaluates encryption of data owner content with uploading time . Data uploading time (DUT) is calculated as:

$$DUT=T\_enc+(T\_end-T\_start) \qquad (1)$$

Where Tenc= total time taken by the method to encrypt the content. Where Tend is data uploading completion time, and Tstartis an initial time of data uploading process.

Data Downloading Time.

In the section, the method proposed defines a mathematical model for downloading time for data in equation (2). In step, ECDPD method computes decryption of data owner content with downloading based on file size. Data downloading time (DDT) is calculated as:

DDT=(T_finished-T_processing)/(File Size)+T_decrypt (2)

Where Tfinished= total time is taken by the method of downloading the content. Where

$$DDT = \frac{T_{finished} - T_{processing}}{File\ Size} + T_{decrypt} \quad (2)$$

Tprocessing is data processing to access & view the content and Tdecryptis decryption time to download the content in original view based on file Size.

Table.1 describes Data Uploading Time (DUT) in a sec and Data Downloading Time (DDT) in milliseconds for 2MB, 4MB and 8MB dataset to perform efficient, portable and secure data de-duplication model in cloud computing environments. In the research work computes the Uploading Time for data (in a sec), and Downloading Time for data (in milliseconds)

along with different length of the dataset. Hence, the research work claims that Proposed ECDPD approach is the best protocol for overall aspects.

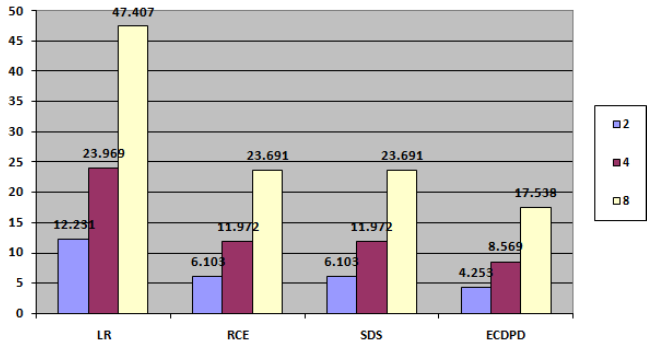

Fig.2 Data Uploading Time for 2MB, 4MB and 8MB dataset Figure 2 displays perform data uploading time (milliseconds) for all existing methods along with proposed efficient content discovery and preserving De-duplication (ECDPD) approach for 2MB, 4MB, and 8MB dataset.

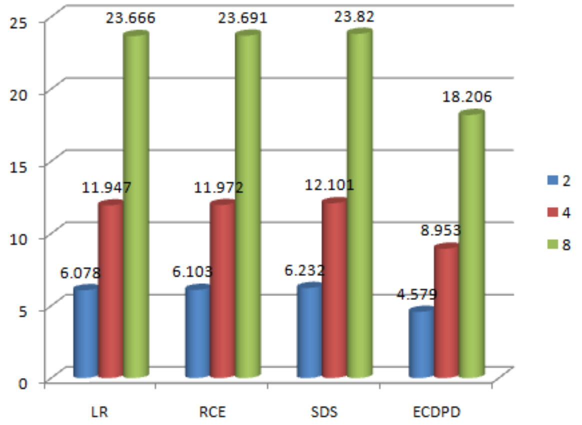

Fig.3 Data downloading time (sec) for 2MB, 4MB and 8MB dataset It Figure 3 presents data transportation time (sec) for all existing methods along with proposed efficient content discovery and preserving De-duplication approach for 2MB, 4MB, and 8MB dataset.

According to Figure 2 and 3 observations, the ECDPD technique is evaluated on DUTandDDT with previous classifiers.The best algorithm for overall data such as 2MB, 4MB, and 8MB data set is ECDPD. The proposed ECDPD is evaluated with LR, RCE and SDS previous mechanisms on behalf of DUT and DDT. RegardingUploading Time for data, RCE, and SDS are the nearest challengers to ECDPD system. Unauthorized data access and updation of tokens are to maintained by RCE. Unauthorized usage and security problems updation with tokens is not permitted by ECDPD. It also fails to maintain less retrieving time. The proposed system maintains less retrieval time. To increase data set size for SDS process, it utilized high retrieval time. The proposed system cannot utilize high retrieval time for changing data set size. Behalf of Downloading Time for data, the LR is the closest existing competitor. But, it cannot be deploy during the data download stage without loss of functionality and effectiveness. The ECDPD can execute the data downloading stage without loss of functionality and effectiveness. It is also enhancing client file range and block range de-duplication with reliability and sharing data files with securely for cloud clients in cloud storage frameworks. The ECDPD reduces 3.802uploading time for data in milliseconds and 3.318downloading time for data in milliseconds. Finally, it is declared that theECDPD algorithm performs best on each evaluation matrix &constraints.

V.CONCLUSION

An efficient content discovery and preserving De-duplication (ECDPD) Algorithm is enhancing client block range and filerange de-duplication with consistency and distributing content owner files with securely for cloud clients in frameworks of storage. The method is protecting the illegal utilization of content owner files accessing and create duplicate copy of content owner file on a cloud storage server to encode the content owner file before storing on cloud storage server. The ECDPD algorithm minimizes 3.802 DUT (Data Uploading Time) in milliseconds and 3.802 DDT (Data Downloading Time). Finally, the paper announces the proposed ECDPD methodology performs best on each estimation matrix & particular input aspects.

REFERENCES

1. Kumaravel A., Meetei O.N.,An application of non-uniform cellular automata for efficient cryptography,2013 IEEE Conference on Information and Communication Technologies, ICT 2013,V-,I-,PP-1200-1205,Y-2013
2. Kumarave A., Rangarajan K.,Routing alogrithm over semi-regular tessellations,2013 IEEE Conference on Information and Communication Technologies, ICT 2013,V-,I-,PP-1180-1184,Y-2013
3. Dutta P., Kumaravel A.,A novel approach to trust based identification of leaders in social networks,Indian Journal of Science and Technology,V-9,I-10,PP--,Y-2016
4. Kumaravel A., Dutta P.,Application of Pca for context selection for collaborative filtering,Middle - East Journal of Scientific Research,V-20,I-1,PP-88-93,Y-2014
5. Kumaravel A., Rangarajan K.,Constructing an automaton for exploring dynamic labyrinths,2012 International Conference on Radar, Communication and Computing, ICRCC 2012,V-,I-,PP-161-165,Y-2012

6. Kumaravel A.,Comparison of two multi-classification approaches for detecting network attacks,World Applied Sciences Journal,V-27,I-11,PP-1461-1465,Y-2013
7. Tariq J., Kumaravel A.,Construction of cellular automata over hexagonal and triangular tessellations for path planning of multi-robots,2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016,V-,I-,PP--,Y-2017
8. Sudha M., Kumaravel A.,Analysis and measurement of wave guides using poisson method,Indonesian Journal of Electrical Engineering and Computer Science,V-8,I-2,PP-546-548,Y-2017
9. Ayyappan G., Nalini C., Kumaravel A.,Various approaches of knowledge transfer in academic social network,International Journal of Engineering and Technology,V-,I-,PP-2791-2794,Y-2017
10. Kaliyamurthie, K.P., Sivaraman, K., Ramesh, S. Imposing patient data privacy in wireless medical sensor networks through homomorphic cryptosystems 2016, Journal of Chemical and Pharmaceutical Sciences .
11. Kaliyamurthie, K.P., Balasubramanian, P.C. An approach to multi secure to historical malformed documents using integer ripple transfiguration 2016 Journal of Chemical and Pharmaceutical Sciences 9
12. A.Sangeetha,C.Nalini,"Semantic Ranking based on keywords extractions in the web", International Journal of Engineering & Technology, 7 (2.6) (2018) 290-292
13. S.V.GayathiriDevi,C.Nalini,N.Kumar,"An efficient software verification using multi-layered software verification tool "International Journal of Engineering & Technology, 7(2.21)2018 454-457
14. C.Nalini,ShwtambariKharabe,"A Comparative Study On Different Techniques Used For Finger – Vein Authentication", International Journal Of Pure And Applied Mathematics, Volume 116 No. 8 2017, 327-333, Issn: 1314-3395
15. M.S. Vivekanandan and Dr. C. Rajabhushanam, "Enabling Privacy Protection and Content Assurance in Geo-Social Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 49-55, April 2018.
16. Dr. C. Rajabhushanam, V. Karthik, and G. Vivek, "Elasticity in Cloud Computing", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 104-111, April 2018.
17. K. Rangaswamy and Dr. C. Rajabhushanamc, "CCN-Based Congestion Control Mechanism In Dynamic Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 117-119, April 2018.
18. Kavitha, R., Nedunchelian, R., "Domain-specific Search engine optimization using healthcare ontology and a neural network backpropagation approach", 2017, Research Journal of Biotechnology, Special Issue 2:157-166
19. Kavitha, G., Kavitha, R., "An analysis to improve throughput of high-power hubs in mobile ad hoc network" , 2016, Journal of Chemical and Pharmaceutical Sciences, Vol-9, Issue-2: 361-363
20. Kavitha, G., Kavitha, R., "Dipping interference to supplement throughput in MANET" , 2016, Journal of Chemical and Pharmaceutical Sciences, Vol-9, Issue-2: 357-360
21. Michael, G., Chandrasekar, A.,"Leader election based malicious detection and response system in MANET using mechanism design approach", Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
22. Michael, G., Chandrasekar, A.,"Modeling of detection of camouflaging worm using epidemic dynamic model and power spectral density", Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
23. Pothumani, S., Sriram, M., Sridhar, J., Arul Selvan, G., Secure mobile agents communication on intranet,Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg No S32-S35, 2016
24. Pothumani, S., Sriram, M., Sridhar , Various schemes for database encryption-a survey, Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg NoS103-S106, 2016
25. Pothumani, S., Sriram, M., Sridhar, A novel economic framework for cloud and grid computing, Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg No S29-S31, 2016
26. Priya, N., Sridhar, J., Sriram, M. "Ecommerce Transaction Security Challenges and Prevention Methods- New Approach" 2016 ,Journal of Chemical and Pharmaceutical Sciences, JCPS Volume 9 Issue 3.page no:S66-S68 .
27. Priya, N.,Sridhar,J.,Sriram, M."Vehicular cloud computing security issues and solutions" Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016
28. Priya, N., Sridhar, J., Sriram, M. "Mobile large data storage security in cloud computing environment-a new approach" JCPS Volume 9 Issue 2. April - June 2016
29. Anuradha.C, Khanna.V, "Improving network performance and security in WSN using decentralized hypothesis testing "Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .

AUTHORS PROFILE

D.Vimala, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India

S. Sangeetha, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India

B. Sundar Raj, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India