A Survey for Secure Encrypted Data Duplication With Dynamic Ownership Updating Using Hash Technique And BVCS (Binocular Visual Cryptography System)

Vignesh Kumar.U[1], Alex Jebaraj J[2], Mohameed Asraff Ali M[3], Ms.Ramya Devi K[4]
M.Tech, Assistant Professor, Dept. of Computer Science and Engineering, S.A. Engineering College, Chennai 600-077

---***---

Abstract: Not with standing industry, the over-burden of data confronting most associations today is a channel on the two people and the venture itself. The expanding volume of this data, which is put away in diverse electronic organizations, requires new modern frameworks to dissect and group them. In this work, we endeavor to actualize a system doca (document classification and analysis) that can disentangle and mechanize such undertakings for various record types, in particular: office archives (content, spreadsheets, and introductions), filtered reports (pictures and pdfs) mixed media records (video and sound). Each record type requires unique strategies for pre-handling, investigation, and characterization. The proficiency and achievability of the doca are inspected on havelsan dataset and precision of various undertakings shows that doca is a promising apparatus for report examination and characterization

1. INTRODUCTION

Cloud storage data model can be used to store data and it is cost efficient in current days, so many duplicate documents are being uploaded easily and securities for the uploaded documents are in great risk. Our project is to eliminate this kind of deduplication and providing security by using hash technique and BVSC . We use Naive hash algorithm to compare the document, to identify the occurrence of deduplication. The security of the uploaded documents maintained by generating two images ,this two images combined and a code is generated to access the document. So unauthorized access are avoided using this method. In Existing System spam mails are filtering on the receiver side. Common solutions include cloud-based mail security products such as Symantec Message Labs and Google Postini, as well as personal security products such as Kaspersky Internet Security and Avast Internet Security.

Mail clients such as Microsoft Outlook and Mozilla Thunderbird, as well as mail service providers, also support spam filtering. The solutions receive mail before filtering, so spamming activities still exist, and spam messages still waste Internet bandwidth and the storage space of mail servers. Spamming bots may access web mail interfaces or deliver via secure SMTP for spamming. Since the packets are encrypted, the detection method cannot identify the spamming bots in this case.

2. Document Classification:

The classification enables more efficient use and protection of critical data. In this modules, the user create new documents can establish processes that enable the user to classify the documents they create, send, modify. They can leave the older data to gradually be retried without being classified. Alternatively, the document classification their backlog of existing data to no modify the current document process of their data.

3. Coding Encryption-Triple DES:

In this modules, the triple DES as an encrypt-decrypt-encrypt process.Authentication process the before sending and receiving data using the system, the receiver and sender identity should be verified. It works by taking three 56-bit keys (k1, k2 and k3) and encryption first with k1, decrypting next with k2 and encrypting a last time with k3.

4. Image Processing BVCS Format:

In this modules, the image processing used by analysis and manipulating the image and output in which result can be altered image or a report which is based on analyzing that image. The BVCS used the model is that it maximize the recovered image in (2, n)-BVCSs. The Hiding of the shared pixel in single image random dot stereogram (SIRDSs) by using an encryption algorithm and a binocular VCS (BVCS) called (2, n) BVCS.

5. Dataset Information:

Dataset is a collection of data, in this module the data are the documents that upload by the users. by using this dataset we can easily verified the document are existing are new document for example, the files will upload by only once if another user going to upload the same document in server they will get the notification (the data is already exist).

6. Reliable Result:

In this modules, the user checks for the file integrity. If the file contains the same word as was in the file previously saved in the server then file will not store instead it shows error to the user. By this the user can have a reliable result for the document they upload.

7. ADVANTAGES:

1) We deal with document diff and similarity analysis for text-based file types, template matching for image files, and content analysis for documents.
2) In the first case, categories are predefined and assigned to the appropriate documents in the training dataset, which is used to train a model that in turn can predict the categories of new documents.

3) In the second case, documents must be clustered automatically without the need for predefined classes.
4) It is in expensive and quick and easy.
5) Less Time and we provide different types of services based on space requirememnts and flexibility.

I. LITERATURE SURVEY

1. **A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data,**Zhihua Xia….,IEEE,

February 2016

- In this document we present multi-keyword ranked search, its dynamically supports the operations like deletion and insertion. Specifically vector and TF IDF model has combined used for query generation and index construction.
- Greedy depth first search provide for efficient multi ranked search and the secure KNN algorithm is used for encrypt and query vectors meanwhile ensure accurate calculation between encrypt and query vectors.

2. **Keyword Search with Access Control over Encrypted Cloud Data,** Zhirong...etal,IEEE, 2016

In this paper, the file can be retrieved only when the keyword match the query and user's attribute value pass the policy check.

KSAC used over encrypted data and HPE used to perform multi-field query search.

3. **D3: A Dynamic Dual-Phase Deduplication Framework for Distributed Primary Storage,**Jianwei Yin…,

IEEE, JULY 2017

In this paper Dynamic dual-phase deduplication framework used maily for storage in distributed manner.

The Dedupe-type used to group data are converted to larger categories,and the coarse-gained filter serves

Inline and offline phase works.

Dynamic performance has been occured by using CTA and DPE mechanisms.

4. **Full Verifiability for Outsourced Decryption in Attribute Based Encryption,** Jiguo Li…etal, *IEEE,* 2017

- In this paper, Attribute based Encryption can be used to reduce the existing decryption cost and limited computing storage space.
- **It checks correctly the unauthorized and authorized transferred cipher text by using the ABE scheme.**

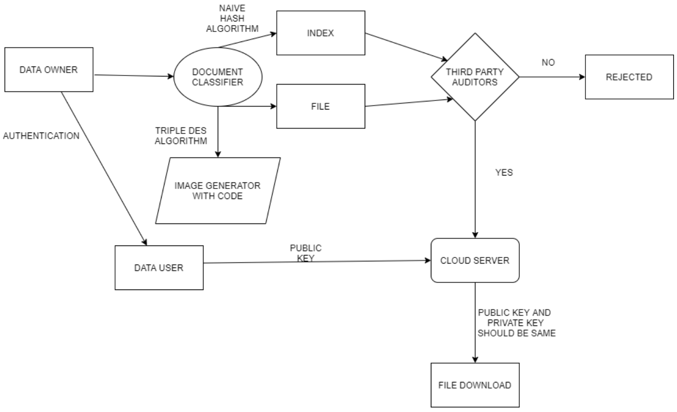
5. **A Fast Asymmetric Extremum Content Defined Chunking Algorithm for Data Deduplication in Backup Storage Systems,**

Yucheng Zhang….,IEEE, 2016

In this paper ,chunk level deduplication plays a role in storage systems, this can be solved by the CDC algorithm called Asymmetric Extremum. Asymmetric Extremum(AE) has been used to resolve the problem of that local range value cannot be replaced by

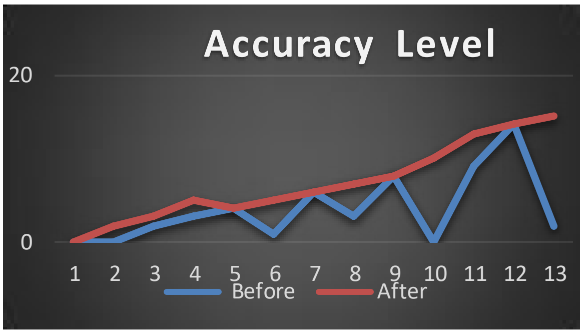
the new extreme range values.

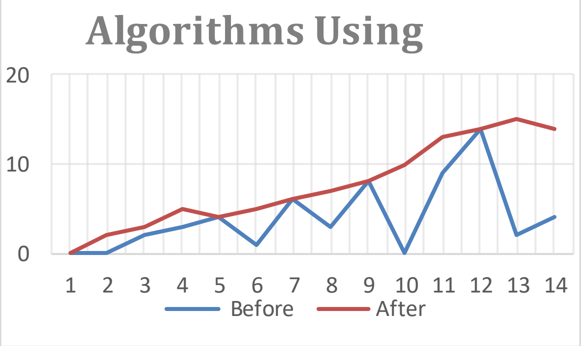
8. Architecture Diagram:



9.CONCLUSION

We accomplished great and dependable outcomes with sensible running times and RAM utilization which makes it a decent structure for general record order and investigation on various document types.It could fill in as a beginning stage for future work on a progressively vigorous, adaptable and include rich arrangement. The entire procedure can likewise be completely mechanized. Future work will incorporate a few enhancements and will address a portion of the present inadequacies: For include coordinating, unique and fresher calculations could be investigated for better execution, For example, FLANN. Something very similar goes for bunching, different strategies like HDBSCAN can be investigated. Just as new vector separation measurements, for example, TSSS. Highlight choice could likewise profit by a few enhancements. For sound substance investigation, unique modules could be prepared for every language since acoustic highlights of speaker sexual orientation are language subordinate.





REFERENCES

[1] X.Yang, R. Lu, J. Shao, et al. "Achieving efficient and privacy-preserving multi-domain big data deduplication in cloud",IEEE Transactions on Services Computing,, 2018.

[2] M.Bellare, S.Keelveedhi, and T.Ristenpart, "Message-Locked Encryption and Secure Deduplication", Advances in Cryptology Eurocrypt 2013, Springer Berlin Heidelberg, 2013, pp.296-312.

[3] M.Bellare, S.Keelveedhi, and T.Ristenpart, "DupLESS: server-aided encryption for deduplicated storage", In Proceedings of the 22nd Usenix conference on Security, Usenix Association, 2013, pp.179-194.

[4] P.Puzio, R.Molva and S.Loureiro, "Cloudedup: Secure deduplication with encrypted data for cloud storage", In IEEE International Conference on Cloud Computing Technology and Science, 2013, pp.363-370

[5] J.Stanek, A.Sorniotti, E.Androulaki, et al, "A secure data deduplication scheme for cloud storage", Ibm Corporation, 2014, PP.99-118.

[6] P.Puzio, R.Molva and M.O¨ nen, "PerfectDedup: Secure Data Deduplication", International Workshop on Data Privacy Management, Springer International Publishing, 2015, PP.150-166.

[7] C.Hui, H.D.Robert and L.Yingjiu, et al,"Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud", IEEE Transactions on Big Data, 2016.

[8] X.R.Ge, J.Yu, H.L Zhang, C.Y.Hu, Z.P.Li, Z.Qin, and R.Hao,"Towards Achieving Keyword Search over Dynamic Encrypted Cloud Data with Symmetric-Key based Verification",IEEE Transactions on Dependable and Secure Computing, 2019.

[9] B.Libert, and D.Vergnaud, "Unidirectional Chosen-Ciphertext Secure Proxy Re-Encryption",IEEE Transactions on Information Theory, 2011, pp. 1786- 1802.

[10] R.Bellafqira, G.Coatrieux and D.Bouslimi,"Proxy Re-Encryption Based on Homomorphic Encryption", Computer Security Applications Conference ACM, 2017, pp. 154-161.

[11] Z.Li, C.Ma, and D.Wang,"Achieving Multi-Hop PRE via Branching Program", IEEE Transactions on Cloud Computing, 2018, doi=10.1109/TCC.2017.2764082.

[12] M.Blaze, G.Bleumer and M.Strauss,"Divertible protocols and atomic proxy cryptography", in Proc. EUROCRYPT 1998, pp.127-144.

[13] Y. Zhou, D. Feng, Y. Hua, et al.,"A similarity-aware encrypted deduplication scheme with flexible access control in the cloud", Future Generation Computer Systems, 2018, 84, pp. 177-189.

[14] B. T. Reddy, T. Rao, "Filter Based Data Deduplication in Cloud Storage using Dynamic Perfect Hash Functions", International Journal of Simulation Systems, Science and Technology, 2018.

[15] J.Li, X.Chen and M.Li, et al, "Secure deduplication with efficient and reliable convergent key management", IEEE transactions on parallel and distributed systems, 2014, 25(6): pp.1615-1625.