JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

De-Duplication Checking in Cloud using Duplication Proof

Mr. Jagati Pavan kalyan (M.C.A). Rajeev Gandhi Memorial college Of Engineering and Technology, Nandyal

***Mr. Dr R. Kaviarasan MTech, Ph.D. Rajeev Gandhi Memorial college Of Engineering and Technology, Nandyal**

Abstract

In particular, the following list summarizes the important contributions of this paper: We suggest a brand-new protocol called TDICP that is based on PIR to verify the accuracy of uploaded files in the CSP when deduplication is being used. TDICP enables users to create unique verification tags for integrity checks, while the verification tags can also, albeit differently, be deduplicated at the CSP. We suggest another cutting-edge protocol called UDDCP to ensure the accuracy of duplication checks based on PSI, making it difficult for the CSP to trick the user into paying for storage space that isn't being used because of deduplication. By reshaping our prior scheme, we create a fresh deduplication scheme called VeriDedup that incorporates the two novel protocols as well as other crucial characteristics, like PoW and data access key assignment, in order to address the deficiencies of its integrity and duplication proof.

1. INTRODUCTION

1.1 Introduction

If the uploaded file has already been saved, verify that it has not been overwritten, and then give the user whose duplicate file data is checked a way to retrieve the file without adding another copy to the cloud. Cloud customers may experience certain security and privacy difficulties, though, as the CSP cannot be completely trusted. Notably, a semi-trusted CSP may alter, falsify, or remove the uploaded data motivated by financial gain. Deduplicated data degradation could result in significant losses for all associated users, such as data owners and holders. Therefore, it is important to check the accuracy of data saved in the cloud, especially when deduplicating duplicate data. It has been suggested that a number of Proof of Retrievability (PoR) approaches be used to alleviate the issue of integrity checks on cloud data storage during the last ten years. In these schemes, a user also adds verification marks to a file. The CSP must use all the information in the user's associated files it stored as inputs to compute a response back to the user during the verification process when the user produces a random challenge and sends it to the CSP. The user then validates the response to ensure the consistency of the stored file. However, current PoR solutions presume that the CSP has limitless computational and storage capabilities and focus primarily on enhancing user performance. Although, the CSP deduplicates data to make the most efficient use of its storage. Regrettably, the above-mentioned existing methods are incompatible with deduplication. This is because distinct verification tags are generated when the same file is owned by various users since the verification tags of these schemes are formed with user specific private keys that are unknown to one another. However, as seen in Fig. 1(a), these verification tags cannot be deduplicated at the CSP. Message-locked In order to verify data integrity when doing deduplication, PoR offers a potential approach. As seen in Fig. 1(b), it uses the message-locked encryption technique to convert a single file into a single verification tag. But such a design.

2. Literature Survey

• **[1] Lal S., Rehman S., Shah J., Meraj T., Rauf H., and Damasˇevičius R:**, Because artificial intelligence (AI) and deep learning (DL) approaches are rapidly expanding, the security and resilience of the deployed algorithms must be ensured. The DL algorithms' vulnerability to adversarial cases has been extensively established. The artificially generated examples will result in many cases that are negatively detected by the DL models but are considered benign by humans. Their characteristics are demonstrated by practical application in genuine physical circumstances with antagonistic dangers. Thus, adversarial attacks and defense, including machine learning and its reliability, have piqued the interest of researchers in recent years. In this section, we offer a system that provides a defensive model against the adversarial speckle-noise attack, adversarial training, and a feature fusion technique that retains classification. And a feature fusion technique that maintains categorization with correct labeling. We assess and analyze adversarial attacks and defenses on retinal fundus images for the Diabetic Retinopathy detection problem, which is regarded as cutting-edge research. Summary: This paper introduces a system that provides a defensive model against an adversarial speckle-noise assault, an adversarial training method, and a feature fusion strategy that preserves classification with proper labeling. We assess and analyze adversarial attacks and defenses on retinal fundus images for the Diabetic Retinopathy detection problem, which is regarded as cutting-edge research.

[2] Rauf H., Lali M., Khan M., Kadry S., Alolaiyan H., Razaq A: The newly discovered human coronavirus illness COVID-19 is the sixth known pandemic following the 1918 flu pandemic. COVID-19 was discovered in Wuhan, China, and quickly spread throughout the world. Almost every country on the planet is facing this natural disaster. We offer forecasting models for the COVID-19 outbreak in Asia Pacific countries, focusing on Pakistan, Afghanistan, India, and Bangladesh. To quantify the severity of the pandemic in the near future, the newest deep learning techniques like as Long Short-Term Memory networks (LSTM), Recurrent Neural Networks (RNN), and Gated Recurrent Units (GRU) are used. When using neural networks, we take into account the time variable and data non-linearity. The key aspects of each model have been analyzed in order to forecast the number of COVID-19 cases in the coming year. 10 days. The predicting performance of the used deep learning models presented up to July 1, 2020, is more than 90% correct, demonstrating the study's dependability. In this study, the most recent deep learning techniques, such as Long Short-Term Memory networks (LSTM), Recurrent Neural Networks (RNN), and Gated Recurrent Units (GRU), are used to estimate the severity of a pandemic soon. When using neural networks, we consider the time variable and data non-linearity. The key aspects of each model have been analyzed in order to forecast the number of COVID-19 cases in the next 10 days. The predicting performance of the used deep learning models presented up to July 1, 2020, is more than 90% correct, demonstrating the study's dependability.

3. OVERVIEW OF THESYSTEM

3.1 Existing System

This model focuses an existing method that is created utilizing some deep learning methods. The technique is carried out here utilizing the ResNet50, which is a transfer learning method, however it does not achieve great accuracy.

3.1.1 Disadvantages of Existing System

- Less feature compatibility
- Low accuracy.

3.2 Proposed System

In our suggested method, we use Convolution Neural Network (CNN) deep learning coupled with CNN transfer learning methods VGG16, CovXNet, and RNN to determine whether a person is sick with pneumonia or not. Pneumonia produces pleural effusion, which is a disease in which fluids fill the lung and cause respiratory problems. Early detection of pneumonia is critical for curative therapy and increasing survival rates. As a result, appropriate classification is essential for the proper therapy that will be feasible with our proposed strategy. The proposed method's block diagram is illustrated below.

3.3 Methodology

In this project work, I used two modules and each module has own functions, such as:

1. Admin Module
2. User Module

User Module

Register Module:

In a distributed system user wants to register with system IP, name and other details which can be stored in databases that can be maintained by admin.

Upload file Module:

In this module we are uploading the file in a database which can be secured by the convergent encryption process, now we are only generating a hash key.

File-level deduplication

Generating hash value for a file and comparing with the hash valueof other files stored in the database.

Block level deduplication

Splitting the files into different blocks. And Generating hash value for each block and comparing with the hash value of other file blocks stored in the database.

View file Module:

This module shows the complete files that can be stored in the database by this User that can be downloaded for person usage.

Admin module

Login Module:

The cloud storage spaces which are maintained by the administrator can authenticate to this module.

Deduplication Module

The duplication is a module which shows the data which tries to upload in the system. So that we can predict which user is supposed to the fake data in the cloud space.

Proof of Ownership

The Ownership of each user which is provided by the owner that is an admin. So that we can avoid duplicates. The admin is which gives single ownership for the single use

4 Architecture

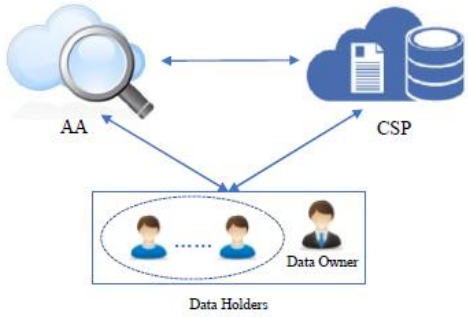


Fig 1: Frame work of proposed method

Algorithm

The scheme proposed is a universal deduplication protocol, meaning which can generally handle all datasets with redundancy, did not specify a dataset in the experiment. Project Uses the hash function SHA-256 to generate a convergent key of size 32 bytes for each block, and also set the size of a tag to be 256 bits for the clearness of comparison. Moreover, we adopt the symmetric-key encryption algorithm Paillier cryptosystem in Cipher-Block Chaining (CBC) mode as the default algorithm to encrypt convergent key

Paillier cryptosystem key generation algorithm

1: Select two large prime numbers p and q where gcd

$(pq,(p-1)(q-1)) = 1$

2: Calculate $n = pq$

3: Calculate $\lambda = lcm(p-1,q-1)$

4: Select g as a random integer where g ∈

5: Define L(x) =

6: Ensure n divides the order of g by checking the existence of the following modular multiplicative inverse

7: u = (L(gλ mod n2))− 1 mod n

8: Public Key = (n,g)

9: Private Key = (λ,u)

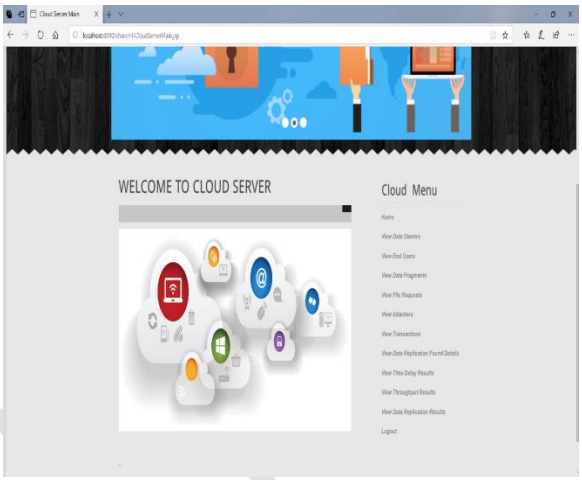
SHA256

The SHA-256 algorithm is one flavor of SHA-256 (Secure Hash Algorithm 256), which was created by the National Security Agency in 2001 as a successor to SHA-1. SHA-256 is a patented cryptographic hash function that outputs a value that is 256 bits long. Secure format that is unreadable unless the recipient has a key. In its encrypted form, the data may be of unlimited size, often just as long as when unencrypted. In hashing, by contrast, data of arbitrary size is mapped to data of fixed size.

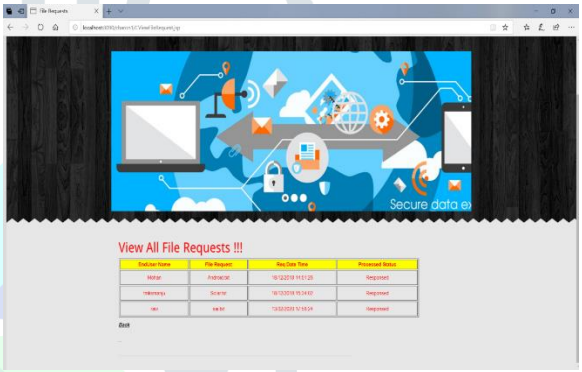
5 RESULTS SCREEN SHOTS

Home Page:

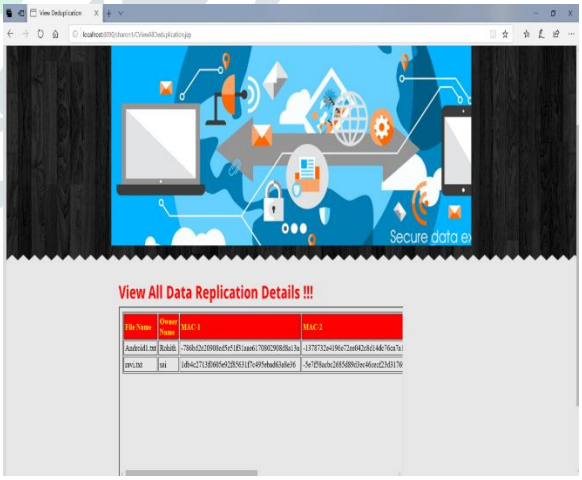
Cloud Menu:



View Files:



Replication Results:



7. CONCLUSION

✓ In this project, we developed VeriDedup to verify the accuracy of the duplication check while also ensuring the integrity of an outsourced encrypted

file. Multiple data holders can independently verify the integrity of their outsourced file using their own unique verification tags using VeriDedup's integrity check protocol TDICP without contacting the data owner. On the other hand, in order to ensure the accuracy of the duplication check, we used a novel challenge and response mechanism in the VeriDedup duplication check protocol UDDCP to let the data holder rather than the CSP first determine whether a file is duplicate. VeriDedup is efficient and secure when used with the security model outlined, according to security and performance analysis. Additionally, the outcomes of our computer simulation.

Future Enhancement

- ✓ In order to safeguard the identity privacy of data owners, which is not considered in our scheme, we will endeavor to find solutions in our future work. In order to safeguard the identity privacy of data owners, which is not considered in our scheme, we will endeavor to find solutions in our future work. The security research demonstrates that our guarantees data confidentiality, convergent key security, and effectively safeguards user ownership privacy all at once. Experimental findings show that our system's security does not come at the expense of its effectiveness. In order to safeguard the identity privacy of data owners, which is not considered in our scheme, we will endeavor to find solutions in our future work.

8. References

[1] Z. Yan, L. F. Zhang, W. X. Ding, and Q. H. Zheng, "Heterogeneousdata storage management with deduplication in cloud computing," IEEETransactions on Big Data, pp. 1–1, 2017.

[2] Z. Yan, W. X. Ding, and H. Q. Zhu, "A scheme to manage encrypteddata storage with deduplication in cloud," in International Conference onAlgorithms and Architectures for Parallel Processing, 2015.

[3] Z. Yan, M. J. Wang, Y. X. Li, and A. V. Vasilakos, "Encrypted datamanagement with deduplication in cloud computing," IEEE Cloud Com-puting, vol. 3, no. 2, pp. 28–35, 2016.

[4] W. Shen, Y. Su, and R. Hao, "Lightweight cloud storage auditing withdeduplication supporting strong privacy protection," IEEE Access, vol. 8,pp. 44 359–44 372, 2020

.[5] Q. Zheng and S. Xu, "Secure and efficient proof of storage withdeduplication," in CODASPY '12, New York, NY, USA, 2012, p. 1–12.

[6] A. Giuseppe, R. Burns, and C. Reza, "Provable data possession at un-trusted stores," in Proceedings of the 14th ACM Conference on Computerand Communications Security, 2007, pp. 598–609.

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, Z. Peterson,and D. Song, "Remote data checking using provable data possession,"ACM Transactions on Information and System Security, vol. 14, pp. 1–34, 2011.

[8] Z. Wen, J. Luo, H. Chen, J. Meng, X. Li, and J. Li, "A verifiable datadeduplication scheme in cloud computing," in INCOS '14, USA, 2014,p. 85–90.

[9] P. Meye, P. Raïpin, F. Tronel, and E. Anceaume, "A secure two-phasedata deduplication scheme," in HPCC '14, CSS '14, ICESS '14, 2014,pp. 802–809.

[10] D. Vasilopoulos, M. Önen, K. Elkhiyaoui, and R. Molva, "Message-locked proofs of retrievability with secure deduplication," in Proceedingsof the 2016 ACM on Cloud Computing Security Workshop, 2016, pp. 73–83.