Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 1, Paper 14 January 2022

Implementation of Data Integrity Evaluation Process in

Dr.R.Poorvadevi[1], M.Tejasri[2] ,N.Ramya[3]
SCSVMV University, Kanchipuram
Tamilnadu, India
rpoorvadevi@kanchiuniv.ac.in

Cloud Computing

***Abstract—DML stands for Distributed Machine Learning. One of the most important AI technologies is DML (distributed machine learning). Nonetheless, in the current distributed machine learning arrangement, data integrity isn't taken into account. However, if network attackers alter or delete the data, the data will be falsified*.As a result, it's vital to maintain data integrity in the DM.It proposes a distributed machine-learned ness initiates data integrity checking approach to verify that the training data is correct.To begin, it recommends using a Provable Data Possession (PDP) selection checking approach to guarantee data integrity. As a consequence, the DML-DIV strategy will be able to withstand manipulation and fraud attacks.Second, as part of the TPA verification procedure, researchers generate a unique integer known as the blinded variable and utilize the discrete logarithm problem (DLP) to build proof and ensure data security.It uses two-step authentication and identity-based cryptographic key generation technology to activate the data possessor's public/private pair of keys.As a result, our DML-DIV method may be able to address the important escrow issue while also cutting document maintenance costs. Lastly,Finally, a thorough theoretical examination is carried out, with formal theoretical analysis and practical evidence confirming the DML-DIV system's safety and efficacy.**

Keywords—cloud platform, Data consistency,Distributed machine learning,Identity-based cryptography,Public auditing.

. INTRODUCTION

Artificial intelligence (AI) has recently been a popular research area in academia and the IT industry.Individuals may use AI to help them with a range of challenges in their everyday life, such as shopping suggestions, transit, facial identification, and self-driving automobilesAs a result, research into artificial intelligence is both theoretically and practically important.

Machine Learning (ML), the foundational technology of artificial intelligence, is the most common method for improving the intelligence of systems and computers. Machine learning is employed in practically every artificial intelligence area.Face recognition, navigation, and self-driving cars, for example, are all feasible thanks to machine learning technologies.

Because of its inefficiency, old machine learning technology can't manage enormous data, especially when the training set is in the petabyte (PB) level or more. To solve the problem, a number of well-known companies, such as Google and Microsoft, have developed Machine learning and artificial intelligence research centres based on huge volumes of data. I'd want to learn more about networked machine learning technology.

1.1 Significance of Distributed Data Integrity Verification Scheme in Cloud Environment using machine learning Approach:

The suggested work is important since an attacker may immediately access our application through URL, and after doing so, it can examine all uploaded files from the data server, as well as edit and save them in the same server's database.

It will also add an analysis graph to the data server, allowing the data server to examine graphs on files that have been attacked and files that have not been attacked. It will introduce a third way, triple DES, to provide greater protection for data stored in cloud computing environments.

1.2 Limitations

Traditional machine learning technology is inefficient when dealing with large amounts of data, especially when the training set is in the petabyte (PB) level or above.

To solve the problem, a number of well-known companies, such as Google and Microsoft, have set up machine learning and artificial intelligence research centers based on huge volumes of data. I'd want to learn more about networked machine learning technologies.The training set in distributed machine learning will be significantly harmed as a result of the prospect of attackers forging, changing, or deleting data.

II. LITERATURE SURVEY

As per authors, S. Hiremath and S. Kunte, "A novel data auditing approach to achieve data privacy and data integrityin cloud computing," explained how the data auditing process is achieved while applying the component of data privacy and integrity feature in the service access platform. [1]

Author S. Jeon et al., "MapReduce Tuning to Improve Distributed Machine Learning Performance”, described the distributed environment machine learning approach to improvise the outcomes in the map-reduce algorithms. [2]

Author W. Hongyuan, "An External Data Integrity Tracking and Verification System for Universal Stream Computing System Framework," has contributed about external data integrity tracking feature and the verification process was determined in the stream computing framework. [3]

III. PROPOSED WORK

In this study, a distributed machine learning literacy approach called Validation of data integrity is developed to assess the consistency of training data in a distributed machine learning system (DML-DIV). In the field of dispensing machine literacy, the DML-DIV system is the first to apply a public selection auditing method to ensure the consistency of the training set.

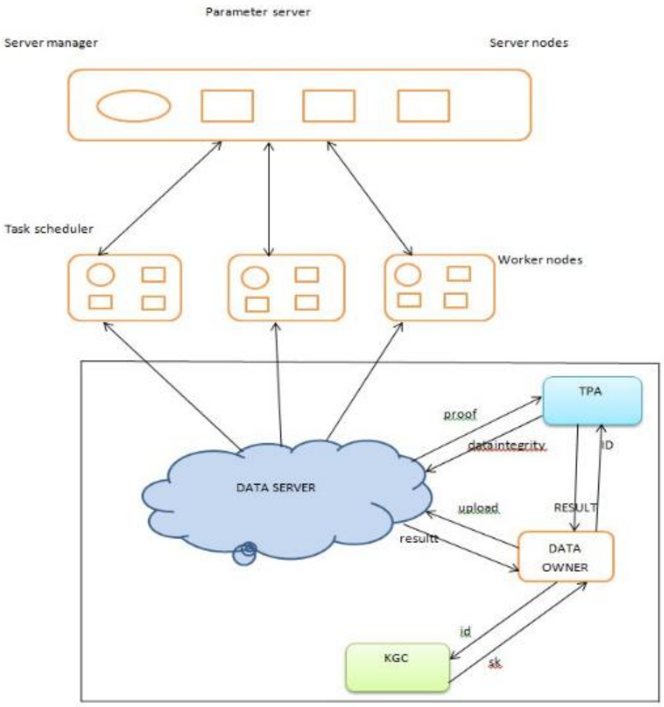


Fig:1.1 Depicting the model of system architecture

The recommended model's functional process is illustrated in the graphic above (figure 1.1), which provides a clear image of the process that occurs in the proposed system.

3.1 Advantages of Proposed Model

- During data integrity audits, addressing the problem of sensitive information sharing.
- Multiple encryptions have been added to the system, making it more powerful.

IV. IMPLEMENTATION WORK

The proposed effort will examine a distributed kind of machine learning technique using powerful mathematical calculations. The following is a list of particular computational requirements and results.

Discrete logarithmic problem (mathematical algorithm)

If Z p and G1 are multiplicative cyclic groups with knowing g, g G1, the benefit of discovering the value is small for any polynomial-timeopponent.

The PPT (Probabilistic Polynomial-Time) Algorithm is a probabilistic polynomial-time algorithm. The PPT (Probabilistic Polynomial-Time) Algorithm is a probabilistic polynomial-time algorithm. Polynomial Probabilistic Probabilistic Polynomial Probabilistic Polynomial Probabilistic Poly When the DLP is successfully solved, the result is AdvDL = Pr[A(g, g) =: Z p], which is unimportant

The probability is calculated using a random selection of on Z p and
a random selection of algorithm A

Encryption Algorithm: RSA

Using the key pair instance technique, create an RSA key. Using the Cipher class, convert an RSA key to encoder format.

Make a key pair with the public and private keys. Using ciphertext c and a public key, compute m such that c me mod n and d e -1 mod ((p-1)(q-1) (n,e).

In encoder format, convert the public and private keys to bytes.

3DES Algorithm

DES encrypts each block twice, instead of just once, using a pair of keys (k1, k2): Ek2(Ek1(plaintext)). If the plaintext is n bits long, this strategy should give equivalent secrecy as using a key length of 2n bits.

A "key bundle" of three 56-bit DES keys (k1,k2,k3) is used in Triple-DES (excluding parity bits). Ciphertext=(Ek3(Dk2(Ek1(Plaintext)) is the encryption algorithm used.

DES encrypts with K1, decrypts with K2, and then encrypts with K3..

- Decryption is the inverse of encryption.:
- Plaintext=Dk1(Ek2(Dk3(Ciphertext))
- In other words, decode with K3, then encode with K2, and then decrypt with K1.

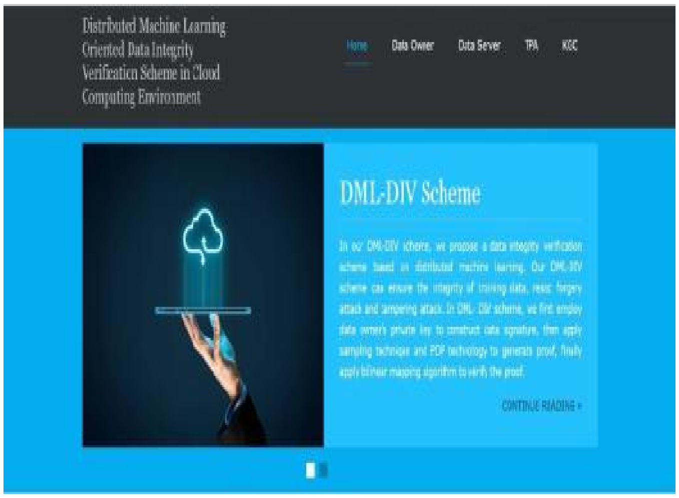
V. *EXECUTION PROCESS AND OUTCOMES*

You can access the home page, data owner, server, key generation center, and third party auditor by login onto the website page with a computer.

The data owner must first create and log in to their account. It then attempts to upload data to the server but is denied until a master key and public key have been established.It is now possible to upload data to the server after creating both keys.

The goal of Kgc (Key Generation Center) is to create a master key and a public key for data security.

All data owners' file blocks, view data, and challenges are visible in the server view.
TPA may access data, audit the data, make challenges, and submit requests to the server for data integrity verification. After the server accepts the request, it can validate the data, and eventually, proofs are delivered to the data owner.


Distributed Machine Learning Oriented Data Integrity Verification Scheme in Cloud Computing Environment
Home
Data Owner
Data Server
TPA
KGC
CONTINUE READING »

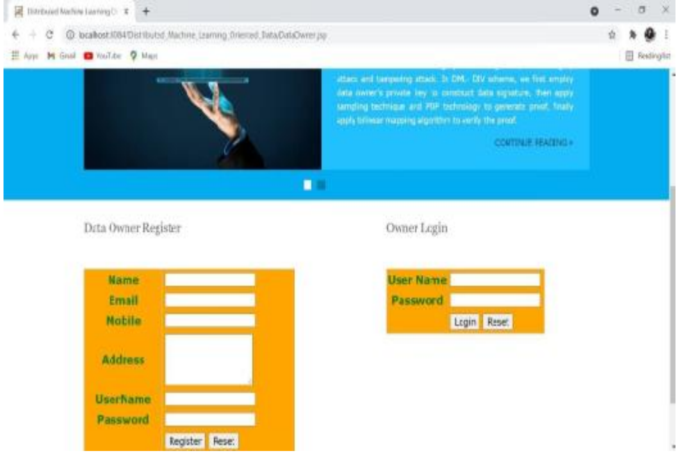
DML-DIV Scheme

In our DML-DIV scheme, we propose a data integrity verification scheme based on distributed machine learning. Our DML-DIV scheme can ensure the integrity of training data, resist forgery attack and tampering attack. In DML- DIV scheme, we first employ data owner's private key to construct data signature, then apply sampling technique and PDP technology to generate proof, finally apply bilinear mapping algorithm to verify the proof.

Figure.1.2 Login Page

Data Owner:

The data owner is in charge of gathering training data and uploading it to the data server. Training data can be collected using computers or mobile devices.


Distributed Machine Learning O
localhost:8084/Distributed_Machine_Learning_Oriented_Data/DataOwner.jsp
Apps
Gmail
YouTube
Maps
Reading list
CONTINUE READING »
Register
Reset
Login
Reset

attack and tampering attack. In DML- DIV scheme, we first employ data owner's private key to construct data signature, then apply sampling technique and PDP technology to generate proof, finally apply bilinear mapping algorithm to verify the proof.

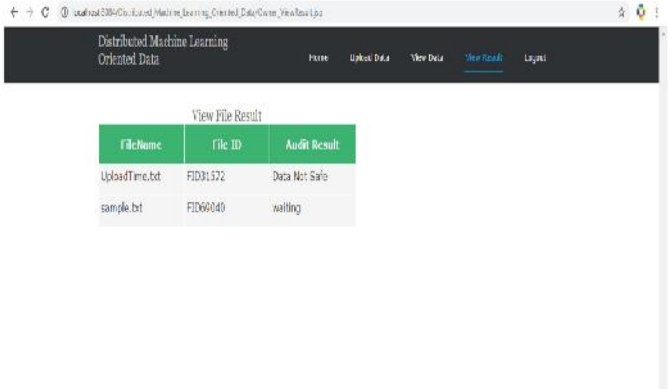
Data Owner Register

Name
Email
Mobile
Address
UserName
Password

Owner Login

User Name
Password

Figure.1.3 Data owner service access credentials



View File Result

FileName	File ID	Audit Result
UploadTime.txt	FID31572	Data Not Safe
sample.txt	FID69040	waiting

Figure.1.4 File selection and integrity checking process

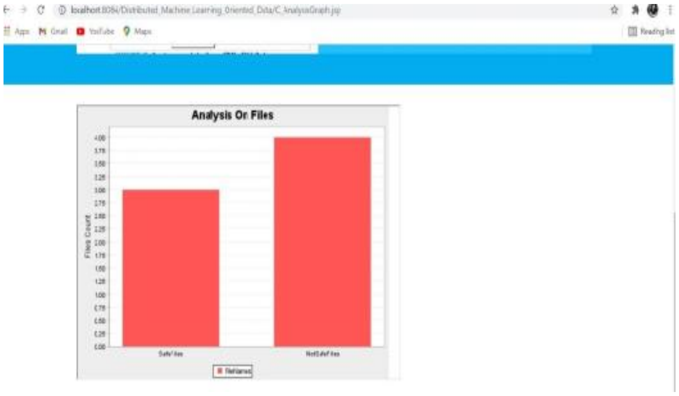
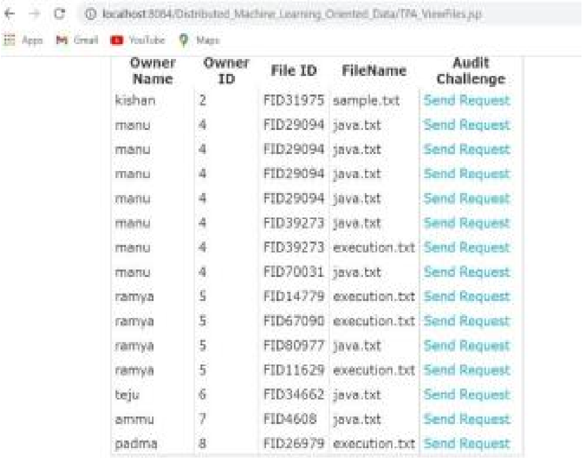


Figure.1.5 Analyzed file information

The integrity feature in heterogeneous platforms is processed in the above diagrams, such as figures 1.3,1.4, and 1.5, and credentials values are validated with the acquired occurrence value, which will be iterated with the performance analysis.



Owner Name	Owner ID	File ID	FileName	Audit Challenge
kishan	2	FID31975	sample.txt	Send Request
manu	4	FID29094	java.txt	Send Request
manu	4	FID29094	java.txt	Send Request
manu	4	FID29094	java.txt	Send Request
manu	4	FID29094	java.txt	Send Request
manu	4	FID39273	java.txt	Send Request
manu	4	FID39273	execution.txt	Send Request
manu	4	FID70031	java.txt	Send Request
ramya	5	FID14779	execution.txt	Send Request
ramya	5	FID67090	execution.txt	Send Request
ramya	5	FID80977	java.txt	Send Request
ramya	5	FID11629	execution.txt	Send Request
teju	6	FID34662	java.txt	Send Request
ammu	7	FID4608	java.txt	Send Request
padma	8	FID26979	execution.txt	Send Request

Figure 1.6 Data Owner Output Page

Key Generation Center:

- The KGC is a body that manages the partial private key of the data owner. To begin, the KGC generates a master secret key and a public key**.** The KGC produces a matching partial secret key for the data owner based on the data owner's ID and master key after receiving the ID from the data owner.

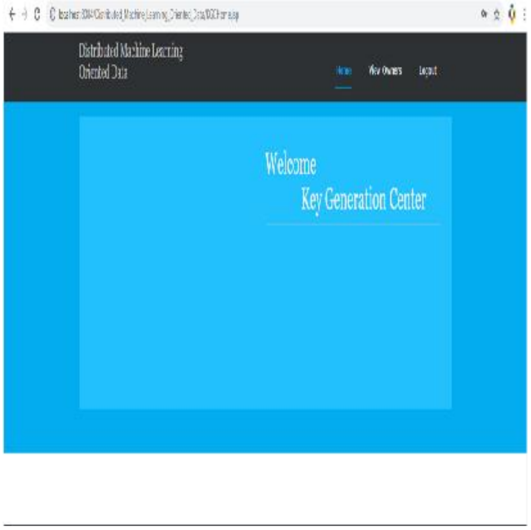


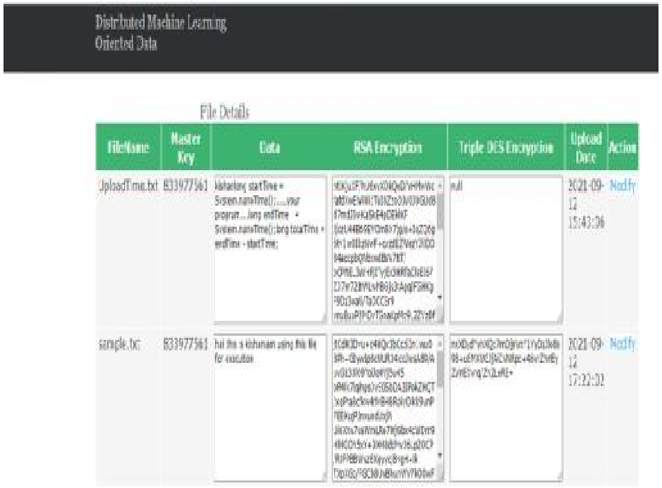
Figure 1.7 Key Generation Center page

Data Server:

The data server is a cloud-based platform that stores the data owner's training data. The data server accepts the TPA's challenge, creates proofs, and delivers proofs back to the TPA. The data server properly executes the protocol and verifies that training data is correctly and thoroughly saved.

Third_Party Auditor:

Third-party auditors can check the consistency of the training set kept on the data server (TPA).The TPA issues a challenge to the data server, and the data server answers with the data owner's master secret key, allowing the TPA to conduct public auditing.


Distributed Machine Learning
Oriented Data

File Details

FileName	Master Key	Data	RSA Encryption	Triple DES Encryption	Upload Date	Action
UploadTime.txt	833977561	kishanlong startTime = System.nanoTime(); ....your program.... long endTime = System.nanoTime(); long totalTime = endTime - startTime;	[illegible]	null	2021-09-12 15:43:06	Modify
sample.txt	833977561	hai this is kishanam using this file for execution	[illegible]	[illegible]	2021-09-12 17:22:32	Modify

Figure 1.7 Attacker details

Network attacker:

By stealing evidence, network attackers attempt to get the data owner's encryption key and divulge the secret of training data. In addition, Network Attackers may use the tempering and forging attacks, as depicted in figure 1.7, to tamper with and counterfeit data proof and identity evidence.

VI. CONCLUSION

This paper proposes a distributed machine knowledge familiartechnique for assuring data integrity in the parameter server setup.TheDML-DIV technology maintains the integrity of the training set storedin the cloud servers by being impermeable to falsification and falseassaults.In addition, our DML-DIV system provides insulation, solvesa major escrow issue, and reduces certificate administration costs.Simulation findings reveal that our DML-DIV system outperformscompeting systems in several circumstances.

ACKNOWLEDGMENT

Implementing the taken problem of distributed data integrity verification scheme was undergone with the various analysis and much thanks to the authors those who are majorly contributed in the same novelty approach.

REFERENCES

[1]. S. Hiremath and S. Kunte, "A novel data auditing approach to achieve data privacy and data integrity in cloud computing," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017, pp. 306-310, doi: 10.1109/ICEECCOT.2017.8284517.

[2]. W. Hongyuan, "An External Data Integrity Tracking and Verification System for Universal Stream Computing System Framework," 2019 21st International Conference on Advanced Communication Technology (ICACT), 2019, pp. 32-37, doi: 10.23919/ICACT.2019.8702046

[3]. H. Zhu et al., "A Secure and Efficient Data Integrity Verification Scheme for Cloud-IoT Based on Short Signature," in IEEE Access, vol. 7, pp. 90036-90044, 2019, doi: 10.1109/ACCESS.2019.2924486

[4]. H. Wang, D. He and S. Tang, "Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity Checking in Public Cloud," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 6, pp. 1165-1176, June 2016, doi: 10.1109/TIFS.2016.2520886.

[5]. H. Wang, D. He, J. Yu and Z. Wang, "Incentive and Unconditionally Anonymous Identity-Based Public Provable Data Possession," in IEEE Transactions on Services Computing, vol. 12, no. 5, pp. 824-835, 1 Sept.-Oct. 2019, doi: 10.1109/TSC.2016.2633260

[6]. Anil Gupta,Dr. Meghna Dubey and Dr. Durgesh Kumar Mishra,"Design& Implementation of Enhanced Security Architecture to Improve Performance of Cloud Computing",Journal of university of shangai for scienceandtechnology,March 2021,vol.23,DOI:10.51201/Jusst12655.

[7].GaopengXie, Yuling Liu, Guojiang Xin, Qiuwei Yang, "Blockchain-Based Cloud Data Integrity Verification Scheme with High Efficiency", Security and Communication Networks, vol. 2021.