ASSURE THE INTEGRITY OF TRAINING DATA USING PUBLIC SAMPLING AUDITING SCHEME

[1]SHAIK HASEENA, [2]D SRINIVAS

[1,2]Dept. of CSE, Kakinada Institute of Engineering & Technology., Matlapalem, Talarevu Mandal, Corangi, E.G.dt, AP, India

ABSTRACT:

In this work, we developed a scheme (DML-DIV) to make sure the reliability of training data. Initially, we agree to the scheme of (PDP) sampling auditing algorithm to get data integrity verification so that our DML-DIV scheme can oppose fake attacks and tampering attacks. Secondly, we generate a random number, namely blinding factor, and apply the discrete logarithm problem (DLP) to construct proof and ensure privacy protection in the TPA verification process. Thirdly, we employ identity-based cryptography and two-step key generation technology to generate data owner's public/private key pair so that our DML-DIV scheme can explain the key escrow problem and reduce the cost of managing the certificates. In conclusion, our analysis outcome shows the security and efficiency of our DML-DIV scheme.

KEYWORDS: DML-DIV(Distributed machine learning oriented data integrity verification scheme), (PDP) Provable Data Possession, TPA, Machine Learning (ML), Artificial Intelligence (AI)

1] INTRODUCTION:

(AI) has turned into a hot examination spot in scholarly world and IT industry lately. Computer based intelligence can assist with tackling different issues in individuals' reality, like shopping suggestion, route, face acknowledgment, and programmed driving. Subsequently, the exploration on man-made brainpower has significant hypothetical worth and down to earth importance.

(ML), as the center innovation of AI, is the crucial method to make PCs and organizations keen. The use of AI traverses all fields of computerized reasoning. For instance, face acknowledgment, route, and programmed driving can be acknowledged through AI innovation.

Because of the helpless productivity of customary AI innovation, this innovation can't manage huge information, particularly when the preparation information arrives at the Peta Byte (PB) level or much bigger. To take care of the issue, some notable organizations, for example, Google and Microsoft have set up enormous information based AI and man-made brainpower research foundations to do the further exploration on appropriated AI innovation. In the interim, the Chinese Computer Society additionally treats the dispersed AI innovation as a significant examination region and the pattern of big data.

2] LITERATURE SURVEY:

2.1] H. Yan, J. Li, J. Han and Y. Zhang *et al*

We give a new proficient RDPC protocol dependent on homomorphic hash work. The new plan is provably secure against falsification assault, supplant assault, and replay assault dependent on an average security model. To help information elements, an operation record table (ORT) is acquainted with track procedure on document blocks. We further give another advanced execution for the ORT, which makes the expense of getting to ORT almost consistent. Besides, we make the extensive exhibition investigation, which shows that our plan enjoys benefits in calculation and correspondence costs. Model execution and investigations display that the plan is achievable for genuine applications.

2.2] H. Zhu, Y Yuan, Y L Chen *et al*

To guarantee data integrity and accessibility in the cloud and IoT storage framework, clients need to check the trustworthiness of remote data. In any case, the current remote DATA integrity check plans are for the most part dependent on the RSA and BLS signature systems. The RSA-based plan has a lot of computational overhead. The BLS signature-based plan needs to embrace a particular hash work, and the bunch signature productivity in the huge information climate is low. Subsequently, for the computational overhead and mark productivity issues of these two mark instruments, we propose a plan of data integrity check dependent on a short signature algorithm (ZSS signature), which upholds security insurance and public evaluating by presenting a (TPA). The computational overhead is adequately diminished by decreasing hash work overhead in the signature process.

3] PROBLEM DEFINTION:

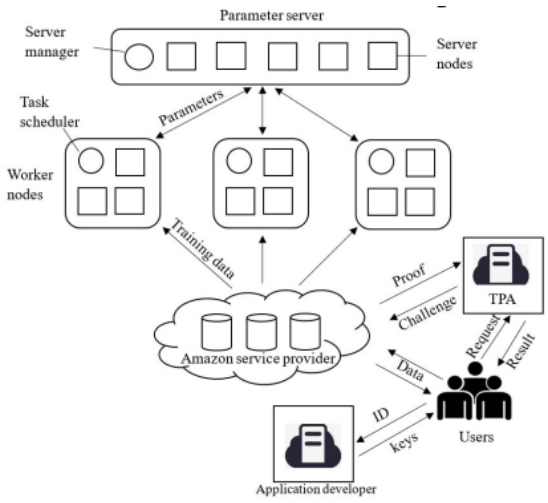
Because of the helpless productivity of conventional AI innovation, this innovation can't manage enormous information, particularly when the preparation information arrives at the Peta Byte (PB) level or much bigger. To tackle the issue, some notable organizations, for

example, Google and Microsoft have set up huge information based ML and AI research establishments to do the further exploration on distributed ML innovation.

4] PROPOSED APPROACH:

To secure the integrity of training data in distributed ML framework, in this work, we propose the distributed ML situated scheme (DML-DIV). As far as we could possibly know, our DML-DIV plot is the main plan in distributed ML region to apply public sampling auditing calculation and guarantee the integrity of training data.

5] SYSTEM ARCHITECTURE:



6] PROPOSED METHODOLOGY:

DATAOWNER

In this application data owner is a module here owner should register with the application and login with application, after successful login he can perform some operations such as upload files, view files, view results and logout.

KGC

Here KGC can directly login with the application and after successful login he can generate public key and master key to owners and at last logout.

TPA

Here TPA can directly login with the application and after successful login he can view all files which are uploaded by owners, after then TPA sends the verification request to data server and after getting the response from data server he can audit the files for checking whether the verified files are got any modification by any attackers or not at last he can sends the verification result to the data owner and logout.

DATA SERVER

Here data server can login with the application here data server can view all owners details, view files details, and view all challenges and logout

7] ALGORITHM:

RSA ALGORITHM:

STEP1: The Public key is used for encryption, and the Private Key is used for decryption.

STEP2: Decryption cannot be done using a public key. The two keys are linked, but the private key cannot be derived from the public key.

STEP3: The public key is well known, but the private key is secret and it is known only to the user who owns the key.

STEP4: It means that everybody can send a message to the user using user's public key. But only the user can decrypt the message using his private key

3-KEY Triple DES:

STEP1: Encrypt the plaintext blocks using single DES with key K_1.

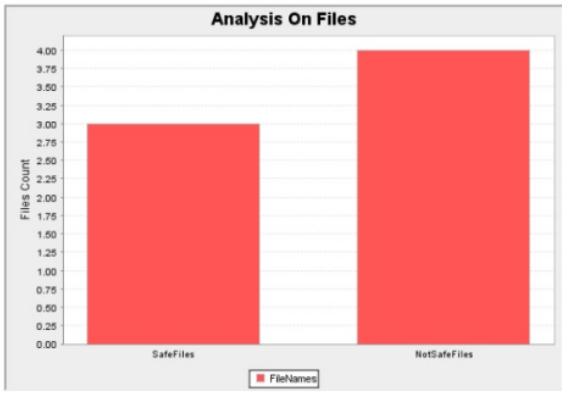
STEP2: Now decrypt the output of step 1 using single DES with key K_2.

STEP3: Finally, encrypt the output of step 2 using single DES with key K_3.

STEP4: The output of step 3 is the ciphertext.

STEP5: Decryption of a ciphertext is a reverse process.

8] RESULTS:



Analysis on file i.e. how many files are safe and how many file are not safe**.**

9] CONCLUSION:

We propose a (DML-DIV) for the parameter server framework. Our DML-DIV scheme can ensure the integrity of the training data stored in the data server, and resist forgery attack and tampering attack. Additionally, our DML-DIV scheme provides privacy protection, solves the key escrow problem, and reduces the cost of managing the certificates. Finally, the simulation results show our DML-DIV scheme performs more efficiently than other schemes.

10] EXTENSION WORK:

We are going to add attackers; Here attacker can directly access our application through URL, after access our page he can view all uploaded files from data server and he can able to modify the files and stores at the same server's database.

And also we adding another facility to data server which is analysis graph, here data server able to view graph on files which are attacked file and which are not attacked files.

In this work we are additionally adding one more algorithm i.e. Triple DES to provide more security to the data which is stored into the cloud computing environment.

11] REFERENCES:

[1] Dean J, Ghemawat S. MapReduce: A Flexible Data Processing Tool. Communications of the Acm, vol. 53, no. 1, pp. 72-77, 2010.

[2] Zaharia M, Chowdhury M, Franklin M et al. Spark: Cluster Computing with Working Sets. 2nd USENIX Conference on Hot Topics in Cloud Computing. 2010, pp. 1-7.

[3] Low Y, Gonzalez J E, Kyrola A, et al. GraphLab: A New Framework for Parallel Machine Learning[J]. Computer Science, vol. 31, no. 1, pp1-4, 2004.

[4] Malewicz G, Austern M H, Bik A J C, et a1. Pregel: a system for large-scale graph processing. Proceedings of the 2010 ACM SIGMOD International Conference Oil Management of data. ACM, 2010, pp. 135-146.

[5] Smola A, Narayanamurthy S. An architecture for parallel topic models. VLDB Endowment, 2010, pp. 703-710.

[6] Dean J, Corrado G S, Monga R, et al. Large Scale Distributed Deep Networks. International Conference on Neural Information Processing Systems. Curran Associates Inc. 2013, pp. 1223-1231.

[7] Douban. Paracel. http://paracel.io/. 2018.

[8] Mu Li, Dave Andersen, Alex Smola. Scaling Distributed Machine Learning with the Parameter Server. International Conference on Big Data Science & Computing. ACM, 2014, pp. 583-598.

[9] Mu Li, Zhou Li, Alex Smola, Parameter server for distributed machine learning. In Big Learning NIPS Workshop, 2013, pp. 1-10.

[10] Ateniese G, Burns R, Curtmola R, et al. Provable data possession at untrusted stores. Acm Conference on Computer & Communications Security. ACM, 2007, pp. 1-25.

[11] Zheng Q, Xu S. Efficient Query Integrity for Outsourced Dynamic Databases. Acm Workshop on Cloud Computing Security Workshop. ACM, 2012, pp. 71-82.

[12] Erway C, Küpçü, Alptekin, Papamanthou C, et al. Dynamic provable data possession. Acm Transactions on Information & System Security, vol. 17, no. 4, pp. 1-29, 2009.

[13] Wang C, Wang Q, Ren K, et al. Privacy-Preserving Public Auditing for Data Storage

Security in Cloud Computing. 2010 Proceedings IEEE INFOCOM. IEEE, 2010, pp. 525-533.

[14] Zhu Y, Hu H, Ahn G J, et al. Efficient audit service outsourcing for data integrity in clouds. Journal of Systems and Software, vol. 85, no. 5, pp. 1083-1095, 2012.

[15] Wang C, Chow S S M, Wang Q, et al. Privacy-Preserving Public Auditing for Secure Cloud Storage. IEEE Transactions on Computers, vol. 62, no. 2, pp. 362-375, 2013.

Ms. SHAIK HASEENA is a student of Kakinada Institute of Engineering & Technology., Matlapalem, Talarevu Mandal, Corangi, E.G.dt, AP, India. Presently she is pursuing her M.Tech[Computer Science and Engineering] from this college and she received her MSC Computer Science from Andhra University College of Engineering, Vishakapatnam in the year 2016. Her area of interest includes Cloud Computing, Machine Learning.

Mr. D. SRINIVAS B.Tech., M.Tech is associate professor in KIET Engineering College. He has 10 years of teaching experience. His area of interest includes Data mining, Networking, Bioinformatics and data science.