ANALYSING THE DATA IN CLOUD AND TO FIND THE FILES WHICH ARE ATTCAKED BASED ON TPA VERIFIED

[1]Mrs. K. MAHALAKSHMI, [2]B.VAISHNAVI
[1]Assistant Professor, [2]UG Research Scholar
[1]Department of B. Com (Business Analytics)

PSGR Krishnammal College for Women, Coimbatore, Tamilnadu, India,

***Abstract*:** The process of analyzing the data in cloud and finding the secured, non-secured files, Encrypted and non-encrypted files and to show that the cloud computing is safe or un-safe for users. The objective of this paper is to find the file which is attacked based on cloud and to analyze whether it is attacked in public or private. Analyzing and visualizing the data and finding the results which are made through Random forest algorithm used for data classification .By using this algorithm finding the accuracy score, weighted average and macro average to show the final results of the privacy cloud dataset.

***Key Words* - Analyzing the cloud data, Random forest algorithm, Public File Attack, Private File Attack.**

I. INTRODUCTION

Cloud computing is a long dreamed vision of computing utility, which enable the sharing of services over the internet. Cloud is a large group of interconnected computers, which is a major change in how we store information and run application. Cloud computing is a shared pool of configurable computing resources, on-demand network access and provisioned by the service provider[1].

By working with cloud server uses virtual platform ESX and a VPN network that links to the client. By simulating different scenarios using the Computer Algebra System Magma tools[2].Cloud Computing a very challenging and potentially formidable task, especially for users with constrained computing resources and capabilities. Thus, enabling public auditability [3] for cloud data storage security is of critical importance so that users can resort to an external audit party to check the integrity of outsourced data when needed.

The purpose of this work is to develop an auditing scheme which is secure, efficient to use and possess the capabilities such as privacy preserving, public auditing, maintaining the data integrity along with confidentiality. To organizations and people cloud provides terribly helpful edges as they\'re alleviated from storage management, investment, and maintenance [4].

Third Party Auditor (TPA) is the secured one, which have the two fundamental requirements have to be met:

1) TPA should be able to efficiently audit the cloud data storage without demanding the local copy of data, and introduce no additional on-line burden to the cloud user;

2) The third party auditing process should bring in no new vulnerabilities towards user data privacy. In this project, we utilize and uniquely combine the public key based holomorphic authenticator with random masking to achieve the privacy-preserving public cloud data auditing system, which meets all above requirements.

To support efficient handling of multiple auditing tasks, we further explore the technique of bilinear aggregate signature to extend our main result into a multi-user setting, where TPA can perform multiple auditing [5] simultaneously. Thus the new auditing scheme has been developed by considering all these requirements. It consist of three entities: data owner, TPA and cloud server. The data owner performs various operations such as splitting the file to blocks, encrypting them, generating a hash value for each, concatenating it and generating a signature on it.

The TPA performs the main role of data integrity check. It performs activities like generating hash value for encrypted blocks received from cloud server, concatenating them and generates signature on it. It later compares both the signatures to verify whether the data stored on cloud is tampered or not. It verifies the integrity of data on demand of the users. The cloud server is used only to save the encrypted blocks of data. This paper provides some right not to be public keeping safe technologies used in cloud computing services.

II. OBJECTIVE

The objective of this project is to find the file which is attacked in cloud and to determine the types of files attacked and analyze whether it is attacked under TPA (Third party auditor) verified files and visualizing whether it is attacked more in public or private data. The random forest algorithm is used in this paper to analyze the data and visualizing the values of the data.

III. RELATED WORK

Privacy concerns arise whenever sensitive data is out- sourced to the cloud. By using encryption, the cloud server (i.e. its administrator) is prevented from learning content within the outsourced databases. But how can we also prevent an area administrator from learning the database content. [6]

Describe a way of securing larger untrusted storage such as a hard-disk. This solution needs a small secure storage for decryption keys. A practical solution based on this approach is the Microsoft Bit locker [7]

while the storage of corporate data on remote servers is not a new development; current expansion of cloud computing justifies a more careful look at its actual consequences involving privacy and confidentiality issues. We introduce in this paper secure privacy preserving cloud database storage architecture. We focus on the Soft- ware as a Service [8]

Describe a technique through usage of a tamper-proof hardware token. This technique applies to a cloud service provider which is based on trusted computing platform. [9]

The data privacy preservation is based on web services, which address the issues of data control and security in a cloud environment. This includes data access, data integrity, data re-regulations [10]

IV. METHODOLOGY

Methodology is the specific procedures or techniques used to identify select, process, and analyze information about a topic. The main goal of this project is to find the files which are attacked under cloud and find whether the files are come under TPA verified and to show it is safe or not for the users to use the cloud computing. In this project Random Forest Algorithm is used to find the result .

4. 1 Random Forest Algorithm

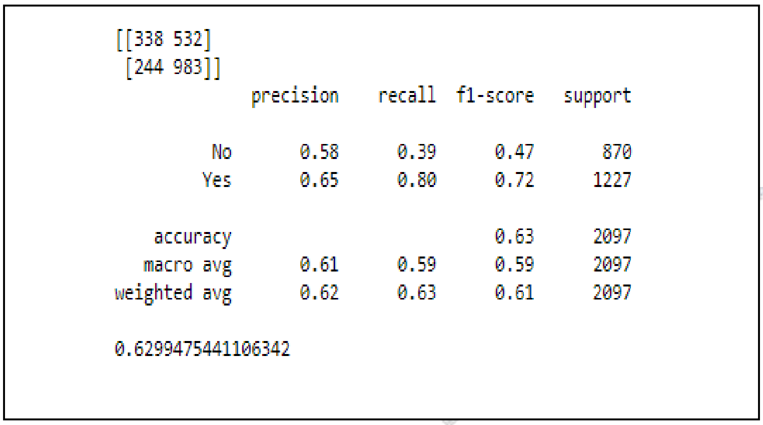
Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. **Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.** Random forest has nearly the same hyper parameters as a decision tree or a bagging classifier. Fortunately, there's no need to combine a decision tree with a bagging classifier because you can easily use the classifier-class of random forest.

Random Forest Algorithm works by completing the following steps

Step 1: select random samples from the dataset.
Step 2: create a decision tree for each sample
Step 3: get a prediction result from each decision tree.
Step 4: perform voting for every predicted result. Use **mode** for classification problem, and **mean** for a regression.
Step 5: select most voted prediction or final prediction.

The above algorithm performing five major steps to finalize the vote prediction. Here, step 1 selecting the samples from the dataset and step2 created decision tree for each decision samples, step3 getting the decision tree predictions to perform voting with classification algorithm assigns **mode** and **mean** for classification and regression for step4. Finally step5, selects most voted prediction.

V. IMPLEMENTATION AND RESULT



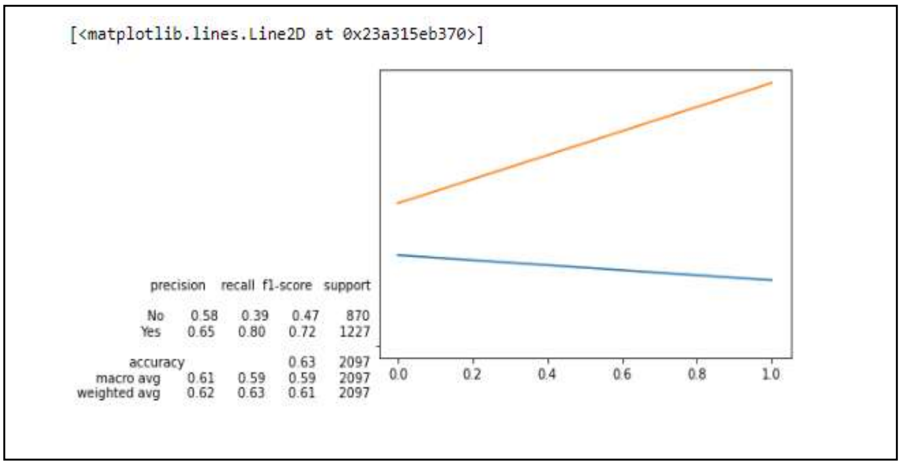
```
[[338 532]
 [244 983]]
              precision    recall  f1-score   support

          No       0.58      0.39      0.47       870
         Yes       0.65      0.80      0.72      1227

    accuracy                           0.63      2097
   macro avg       0.61      0.59      0.59      2097
weighted avg       0.62      0.63      0.61      2097

0.6299475441106342
```

The above figure explains the train split function in order to make the split and giving the values to the variables for x_train and y_train (prediction variable and TPA verified) and fitting the model on the training data and trying to predict the attributes. To find the file which is attacked based on cloud and to analyze whether it is attacked in public ,private or hybrid. It predicted the values of the accuracy score ,macro average and weighted average as same i.e.,2097.and matplotlib is imported pyplot in order to plot graphs of the data.



```
[<matplotlib.lines.Line2D at 0x23a315eb370>]

              precision    recall  f1-score   support

          No       0.58      0.39      0.47       870
         Yes       0.65      0.80      0.72      1227

    accuracy                           0.63      2097
   macro avg       0.61      0.59      0.59      2097
weighted avg       0.62      0.63      0.61      2097
```

The above figure is to find the files that are verified by TPA and the files which are affected and the types of files have been affected under public and private. And it shows the result that is there are more number of encrypted files than a non-encrypted file.so it states that the more files are affected under public cloud computing.

VI. CONCLUSION

The main aim of this project is to find the attacked files in cloud and whether they are attacked under TPA verified files and types of files attacked under public and private cloud files. And to show whether the cloud computing is safe for users or not .By taking the random forest algorithm in this project we analyzed and visualized the data and finded the accuracy, macro average and weighted average of the attribute.

REFERENCES

[1] Mell P, Grance T. Draft NIST working definition of cloud computing.

[2]WiebBosma, John Cannon, and Catherine Playoust. The Magma algebra system I: The user language. J. Symbolic Comput., 24(3-4): 235-265, 1997. Computational algebra and number theory, London, 1993.

[3]Qian Wang and Cong Wang and Kui Ren, Wenjing Lou, Jin Li "Enabling Public Auditability And Data Dynamics For Storage Security in Cloud Computing" in IEEE transactions on parallel and distributed systems, 2011, vol. 22, no. 5.

[4]To organizations and people cloud provides terribly helpful edges as they\'re alleviated from storage management, investment, and maintenance

[5]Ravi Kant Sahu and Abhishek Mohta, L.K. Awasthi "Robust Data Integration While Using Third Party Auditor For Cloud Data Storage Services", conf. IJARCSSE, 2012, Volume 2, Issue 2,ISSN: 2277 128X.

[6]Hacigu¨mu¨s¸, B.Iyer, and S.Mehrotra, "Providingdatabase as a service," in Proceedings of the International Conference onDataEngineering,LosAlamitos,CA,USA,2002.

[7]Maheshwari, R. Vingralek, and W. Shapiro, "How to create a trusted database system on untrusted storage," in Proceedings of the 4th USENIX Symposium on OS Design and Implementation, Berkeley, CA, USA, 202000.

[8] Ferguson, "AES-CBC+ Elephant diffuser A Disk .Encryption Algorithm for Windows Vista," Microsoft, 2006.[Online].Available: http://download.microsoft.com/download/0/2/3/0238acaf d3bf-4a6d-b3d6-0a0be4bbb36e/bitlockercipher200608.pdf

[9] Mowbray, S. Pearson, and Y. Shen, "Enhancing privacy in cloud computing via policy-based obfuscation." Springer Berlin/Heidelberg,2010,pp.1–25.

[10] David C,Fatema K (2012) A privacy preserving authorisation system for the cloud.J Comput Syst Sci 78(5):1359-1373.