PRIVACY PRESERVATION IN CLOUD STORAGE BASED ON CLOUD TYPE

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

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

Abstract Analyze the number of files being attacked, in what technique did the files being attacked and what type of files were not attacked. In cloud computing models are filled with advantages compared to on site models, they're still vulnerable to both inside and out of doors attacks. Therefore, cloud developers got to take security measures to guard their users sensitive data from cyber attacks. While using the Random forest classifier algorithm analyze and visualizing the data and finding the cloud type and attacks within the cloud computing infrastructure aren't new ,but attacks supported the duplication feature within the cloud computing is comparatively new and has made its urge nowadays.

Key Words- Cloud type, Random forest classifier algorithm.

1. INTRODUCTION

Cloud computing is that the on-demand availability of computing system resources, especially data storage (cloud storage) and computing power, without direct active management by the user. The term is typically used to describe data centers available to many users over the online. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is comparatively close, it's going to be designated a foothold server. Cloud computing is dramatically changing the way that organizations manage their data, due toitsattractivefeatureslikerobustness,lowcost,andubiquitousnature.However,privacyconcernsarise whenever sensitive data is outsourced to the cloud where the info is processed and stored. The fact that users not have physical possession of the outsourced data makes it a formidable task to understand the data confidentiality and integrity. As the data, in most cases encrypted, need to be not only stored, but also processed in clouds, the cryptography-based data confidentiality and integrity protection approaches aren't adequate to satisfy the security requirements. Privacy preserving in cloud environments includes two aspect: processing Data processing security covers the problems of the way to protect user privacy time during a virtualized cloud platform. Data storage security covers the issues of guaranteeing user data privacy when the data is stored in data center. This special issue consists of eight papers addressing the safety and privacy issues in cloud computing[1][2][3]

2. OBJECTIVE

To analyze the number of files being attacked and what type of files were attacked and TPA method and Third Party Auditing (TPA) give protection to the files. Cloud developers need to take security measures to protect their user's sensitive data from cyber attacks.

3. RELATED WORK

The increasing volume of personal and sensitive data being harvested by data controllers make it increasingly necessary to use the cloud not just to store the data, but as to process them on cloud premises. However, security concerns on frequent data breaches, along side recently upgraded legal data protection requirements advise against outsourcing unprotected sensitive data to public clouds. To tackle this issue, this survey covers technologies that allow privacy-aware outsourcing of storage and processing of sensitive data to Public cloud [5]

Large scale distributed systems especially cloud and mobile cloud deployments provide great services improving people's quality of life and organizational efficiency. cloud computing engages with the perils of peer-to-peer (P2P) computing and brings up the P2P cloud systems as an extension for federated cloud. According to the experimental power and delay results, the hybrid cloud model performs up to 75% better as compared to the traditional cloud models. [4]

The existing technologies and a good array of both earlier and state-of-the-art projects on cloud security and privacy. We categorize the prevailing research consistent with the cloud reference architecture orchestration, resource control, physical resource, and cloud service management layers, additionally to reviewing the prevailing developments in privacy-preserving sensitive data approaches in cloud computing like privacy threat modeling and privacy enhancing protocols and solutions.[7]

Location-based services(LBSs) are increasingly popular in today's society. People reveal their location information to LBS providers to obtain personalized services such as map directions, restaurant recommendations, and taxi reservations. This paper contains security analysis and performance experiments to demonstrate the privacy-preserving properties and efficiency of our proposed scheme.[8]

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 in the outsourced databases. But how can we also prevent a local administrator from learning the database content. And how can we avoid scenarios such as: employees using cloud applications may learn more than it is necessary to perform their respective duties.

Random forests is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting.

Random forests has a variety of applications, such as recommendation engines, image classification and feature selection. It can be used to classify loyal loan applicants, identify fraudulent activity and predict diseases. It lies at the base of the Boruta algorithm, which selects important features in a dataset.

Random forest uses gini importance or mean decrease in impurity (MDI) to calculate the importance of each feature. Gini importance is also known as the total decrease in node impurity. This is how much the model fit or accuracy decreases when you drop a variable. The larger the decrease, the more significant the variable is. Here, the mean decrease is a significant parameter for variable selection. The Gini index can describe the overall explanatory power of the variables.[10]

4. METHODOLOGY

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a mess of decision trees attaining time and outputting the category that's the mode of the classes(classification)or mean/average prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forests generally outperform decision trees, but their accuracy is less than gradient boosted trees. However, data characteristics can affecttheirperformance.Decisiontreesareawell-likedmethodforvariousmachinelearningtasks.Treelearning"come[s]closesttomeeting the wants for serving as an off-the-shelf procedure for processing However, they're seldom accurate".[9]

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 file attack method 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 **mod e** for classification problem, and **mean** for a regression.
Step 5: select most voted prediction or final prediction.

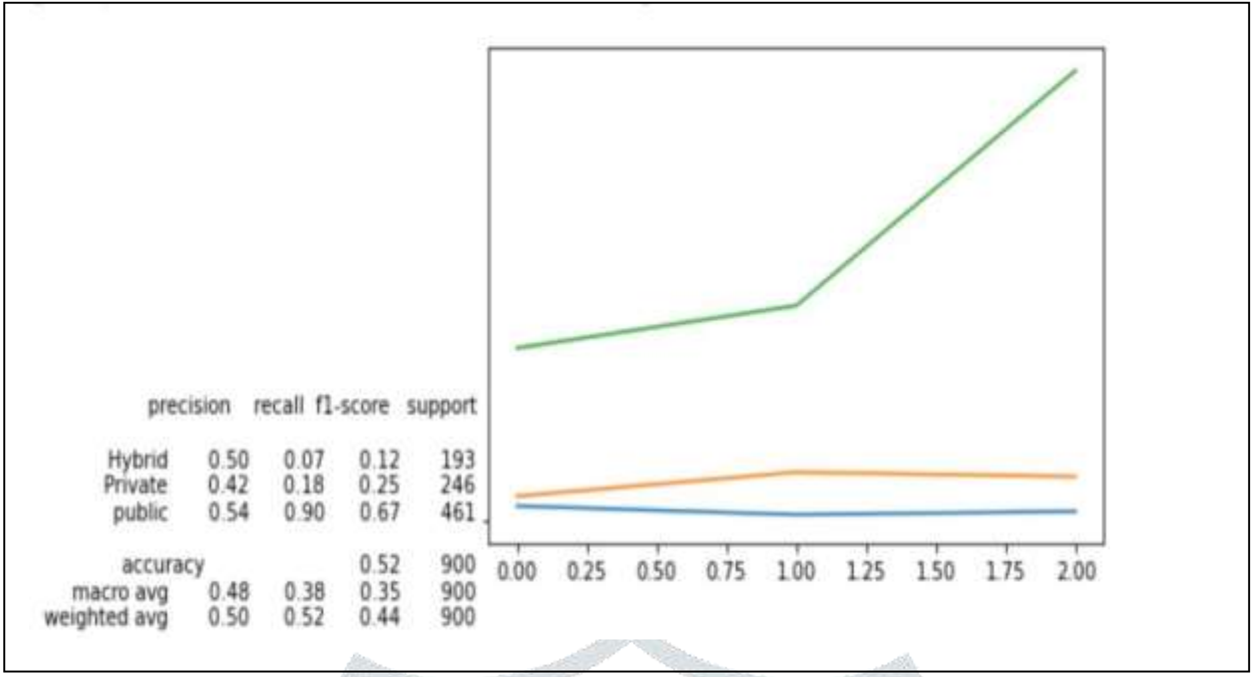
The above algorithm performing five major steps to finalize the vote prediction. Here, step 1 selecting the samples from the dataset and step2 created decision tree for each decision samples, step3 getting the decision tree predictions to perform voting with classification algorithm assigns **mode**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

5. RESULTS AND DISCUSSION

Accuracy			0.52	900
Macro avg	0.48	0.38	0.35	900
Weighted avg	0.50	0.52	0.44	900

Fig5.1 predicted value

The Above fig5.1 Diagram Represents The cloud type, what types of files being attacked by using cloud type attribute to analyze the full accuracy of ‘cloud type'. Which one is highly secured data. As per the analysis of accuracy Macro average and weighted average analysis are same and the values were near 900, 900, 900 repeatedly. Macro average is less than Weighted average.



```
              precision    recall  f1-score   support

      Hybrid       0.50      0.07      0.12       193
     Private       0.42      0.18      0.25       246
      public       0.54      0.90      0.67       461

    accuracy                           0.52       900
   macro avg       0.48      0.38      0.35       900
weighted avg       0.50      0.52      0.44       900
```

Fig5.2 Cloud Type

The above fig5.2 diagram represents the cloud type and number of files and what types of files being attacked by using cloud type attribute to analyze the full accuracy of ‘cloud type'. Which one is highly secured data. In public support value is 461 it means highly attacked and Hybrid support value is 193 it means less attacked so therefore Hybrid is secured more than public and private. As per the analysis of accuracy Macro average and weighted average analysis are same and the values were near 900, 900, 900 repeatedly. As the analysis shows that as the total attack in the cloud type attack affected by all were 0.4711111111111111

VI. CONCLUSION

The project is to find out the privacy preservation in cloud storage based on cloud type, here we have predicted cloud type, TPA verified. The security and privacy in cloud computing is neglected, then the privacy information of each user is at risk .Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust, and every phase of design. The verification process used in this mechanism is able to support hybrid authentication protocols.

References

1. Ho, Tin Kam(1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 5 June2016
2. Jumpupto:**abcd**HoTK(1998)."TheRandomSubspaceMethodforConstructingDecisionForests"(PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence. **20** (8): 832–844.doi:10.1109/34.709601.
3. Jumpupto:**abcdefg**Hastie,Trevor;Tibshirani,Robert;Friedman,Jerome(2008).The Elements of Statistical Learning(2nded.). Springer. ISBN0-387-95284-5
4. A. AlDairi, L.Tawalbeh

   Cyber security attacks on smart cities and associated mobile
   technologies Proc. Comput. Sci., 109 (2017), pp. 1086-1091
5. Josep Domingo-Ferrer, OriolFarras, Jordi Ribes- Gonzalezand DavidSanchez
6. S. Pearson, “Privacy, security and trust in cloud computing,” in Privacy and Security for Cloud Computing (S. Pearson and G. Yee, eds.), Computer Communications and Networks, pp. 3–42, Springer London,2013.
7. Statista, “Number of location-based service users in the United States from 2013 to 2018 (in millions),” Statista;2017.
8. Maheshwari, R. Vingralek, and W. Shapiro, “How to build a trusted database system on untrusted storage,” in Proceedings of the 4th USENIX Symposium on Operating System Design and Implementation, Berkeley, CA, USA,202000.
9. https://www.datacamp.com/community/tutorials/random-forests-classifier-python.