# Cloud Based Reasoning Health Data using Homomorphic Encryption

**S.Devi, S.Poornima, Kaviya Sri A.M, Monisha Devi .G, Mownika. M**

***Abstract*: Cloud computing is an abundant heterogeneous paradigm. The clients are given access to cloud for storing large amount of data for many purposes. The major cloud security issues are data breaches, insider threat and insufficient due diligence etc. Most of the service providers save the Client data as a plain text format which makes the data less secured. Aim of the system is to protect the health data that are outsourced for storing in cloud. In this system, the data is encrypted using paillier cryptosystem before outsourcing, which preserves the privacy of patient's health data. Computations are performed over this encrypted data using decision tree algorithm. The results are displayed on the client machine. Hence, it ensures the privacy preservation and cautions the patient about his health.***

***Keywords: homomorphic encryption, machine learning, Opennebula, private cloud***

## I. INTRODUCTION

For decades people have been storing the data in computer's hard disk, flash media, CD-ROM such other devices and access it whenever needed. As the technology has been growing drastically the need to store things in hard disk no longer needed. Here is where cloud comes into picture. To store our information we no longer need hard disk. Cloud can store enormous amount of data where we can use it through internet from anywhere in the world.

To store such a huge creating a physical infrastructure is time consuming and expensive. It would be easier if we can store in an infrastructure that is already exists which can be accessed through a common source. Here is where cloud computing comes into play. Cloud computing is merely based on resource sharing.

One of the most influential and powerful technologies in today's world is Machine Learning. One of the most widely used and practical methods of supervised learning is Decision Tree. It is method used for both classification and regression tasks. An algorithmic approach is constructed that identifies ways the data set split based on different conditions. A model is created and the value of a target is predicted by learning decision rules inferred from the data is necessary. The rules are generally in form of if-then-else statements. Deeper the tree, the more complex the rules and fitter the model.

## II. LITERATURE SURVEY

In cloud computing, one of the main concerns of the users/clients is its security. Since we are computing patient data it is still more confidential. When the confidentiality is lost then there is no privacy preservation. Hence it is highly required to secure its data. So as the patient data is maintained with integrity and at most confidentiality.

Xiaoliang Wanga et al.(2019)[1] implemented a new method to solve a problem of processing data in cloud center. He proposed a schema where eHealth data uses fully homomorphic concept for privacy preservation. So that the data contents are not revealed once its encrypted.

Tessema Mengistu et al.(2017)[2] proposed CuCloud. Cu Cloud uses unused resources of underutilized computer within an organization or community. There is no need of data center when CuCloud is established in any organisation.

Suneeta Mohanty et al.(2018)[3] proposed that Cloud Computing Environment helps users of the cloud to store their data in them which low down the burden of data storage. In this data privacy is maintained and it also includes integrity, confidentiality and data availability.

Mai Rady et al. (2019)[4] implemented TPA authentication and non repudiation and also proposed Auditing scheme for achieving integrity, confidentiality.

Rajat Saxena et al.(2017) [5] proposed a effective scheme that is also experimentally tested which uses Hadoop and MapReduce framework and it is tested against various parameters.

Shariqua Izhar et al.(2017)[6] proposed a algorithm which is most based on various cipher algorithm and symmetric key encryption in which it also contains concept of confusion and diffusion.

N. Leavitt et al.(2009)[7] proposed that there is a bright future for cloud computing though it faces several significant challenges.

A. Rao (2012) [8] proposed a RSA key exchange protocol between cloud service provider and the user for secret sharing of symmetric key.

## III .PROPOSED SYSTEM

Firstly, a Private Cloud is setup using Opennebula. This cloud has a server and a client. Where the client is used as a user interface by the patients. Client side collects the information likes blood pressure, maximum heart rate, serum cholesterol, chest pain type, electrocardiographic results, exercise induced angina. This itemset is sent to the Server Side when these data are firstly encrypted using Paillier cryptosystem.

The encrypted values are stored for futuristic use and also sent for prediction of disease. The prediction of the disease is done using decision tree algorithm. Decision tree algorithm takes one entity as a root node and predicts others.

Revised Manuscript Received on April 15, 2020.

**S. Devi***, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, India.devi.s@cit.edu.in

**S. Poornima**, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, India.poornima.s@cit.edu.in

**Kaviya Sri.A.M**, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, India. Kavyaam31@gmail.com

**Monisha Devi.G**, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, India.g.monishadevi2k@gmail.com

**Mownika. M**, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, India.mownika21@gmail.com

The predicted results are sent to client side and displayed to the users. The below figure (Fig.1.1) depicts the system architecture. The program has been implemented using *django* framework. The Client Side requests the database (Server) for the prediction. Where in the database analyzes the given input and predicts the disease. The results are displayed in the Client Side.

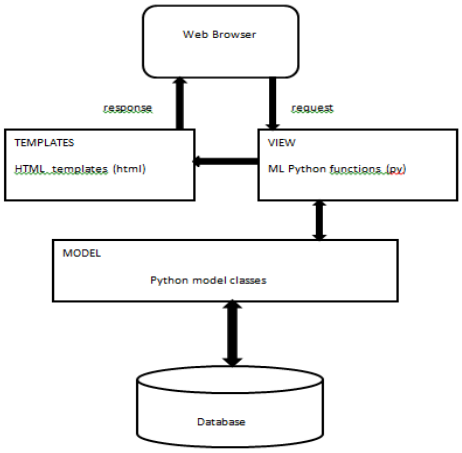


**Fig.1.1. System Architecture.**

## IV. METHODOLGY

Encryption of the patient's details is done using paillier encryption. Paillier is a type of Key pair-based cryptography. The item set given by the user is encrypted so that the patient's data is not leaked. In paillier we choose two large prime independent number such that the greatest common divisor equals to 1.

Then we are urged to find the lambda. Where the lambda equals to the prime independent numbers minus 1.

Select random integer g. We name the multiplication of the two independent numbers as n. where the modular multiplicative inverse is calculated.

The public key consists set of n & g and private key is the set consisting of lambda and values calculated for modular multiplicative inverse.

Encryption:

1. Let the message to be encrypted be m, where m should be greater than or equal to zero but lesser than n.
2. Select random r where r is greater than zero and lesser than n. Ensure that greatest common divisor between r and n equals to 1.
3. Compute cipher text.

There is no need for to decrypt the text at any point of the time because we are computing it upon the encrypted text.

After the items are encrypted it is stored in the database. The encrypted data is further taken for prediction using decision tree algorithm. Predicting or operations upon an encrypted text is known as Homomorphic Encryption.

One the popular classification algorithm is the decision tree algorithm. There are two steps involved in classification they are learning step and prediction step. In learning step, using the training data a model is developed. In prediction step, this model is used to predict the response for given data.

On the encrypted text, calculate the target entropy. Then the dataset is split based on various attributes. For each branch the entropy is calculated and added proportionally to get the total entropy for different attributes. The resulting entropy is subtracted from the entropy before split. The result obtained is the Information Gain or decrease in entropy. The largest information gain as the decision node is chosen as an attribute, divide the dataset by its branches and repeat the process on every branch. These operations are done upon an encrypted text without decrypting it is called homomorphic encryption.

## V. RESULT ANALYSIS

Item set is taken as an input from the user. The item set consists of Age, Gender (1=Male;0=Female), Fasting Blood Sugar (1=Yes; 0=No), Blood Pressure, Serum Cholesterol, Maximum Heart Rate, Chest Pain Type, Resting Electrocardiographic results, exercise induced angina etc.

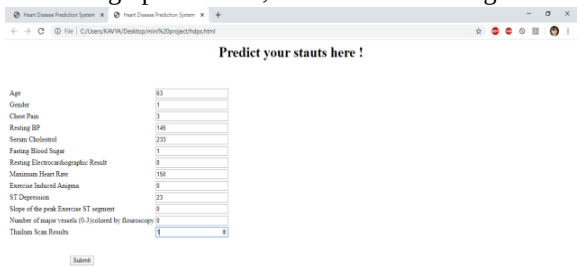


**Fig 5.1. ItemSet**

These item sets are encrypted and the encrypted data is being sent for prediction. The prediction is done using decision tree algorithm. (Fig 5.2) The result is displayed to the user. If results is 1= Disease and result is 0= No Disease. The prediction Status will inform us the whether we are predicted with heart disease or not.


**Fig 5.2. Prediction Results**

## VI. FUTURE SCOPE AND CONCLUSION

The experiment is conducted with different item-sets and the results have been evaluated to be 90% accurate. This paper mainly focuses on securing health data other issues like scheduling, making it cost effective and avoiding server crash or using an algorithm which has less space and time complexity can be taken as the future scope.

## REFERENCES


1 Xiaoliang Wanga, Liang Bai, Qing Yanga, Liu Wanga, Frank Jianga, "A dual privacy-preservation scheme for cloud-based eHealth systems", Elsevier - Journal of Information Security and Applications, Vol. 47 (2019), 132–138, 2019.

2 Tessema Mengistu, Abdulrahman Alahmadi, Abdullah Albuali, Yousef Alsenani, and Dunren Che, "A 'No To Data Center' Solution to Cloud Computing", IEEE 10th International conference, 2017.

3 Suneeta Mohanty, Prasant Kumar Pattnaik, Raghvendra Kumar, "Confidentiality Preserving Auditing for Cloud Computing Environment", IEEE 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), 2018.

4 Mai Rady, Tamer Abdelkader, Rasha Ismail, "Integrity and Confidentiality in Cloud Outsourced Data", Ain Shams Engineering Journal, Vol. 10 (2019).

**5** Rajat Saxena, Somnath Dey, "Cloud Audit: A Data Integrity Verification Approach for Cloud Computing", international multi-conference on Information processing -2017.

6 Shariqua Izhar, Anchal Kaushal, Ramsha Fatima, Mohammed A. Qadeer, "Enhancement in Data Security using Cryptography and Compression", 7th International Conference on Communication Systems and Network Technologies, 2017.

7 N. Leavitt, "Is cloud computing really ready for prime time?" Computer, vol.42, no.1, pp.15–25,2009.

8 A. Rao, "Centralized database security in cloud," International Journal of Advanced Research in Computer and Communication Engineering, vol.1, pp.544–549,2012.


## AUTHORS PROFILE

**Devi S** received her B.E degree in Computer Science and Engineering,M.E degree in Computer Science Engineering from Anna University. She has 6 years of teaching experience. Her area of research includes Big Data Analytics, Security issues in Data Analysis, Natural Language Processing.

**Poornima S** received her Master degree in information Technology from MIT Campus, Anna University, Chennai. Her area of research includes Big data analytics, Video analytics and Speech processing

**Kaviya Sri A.M.** an aspiring undergraduate student in Information Technology from Coimbatore Institute of Technology, Coimbatore. Completed a Course in Cloud computing, doing research in the area of cryptography.

**Monisha Devi G** Studying B.Tech Information Technology in Coimbatore Institute of Technology,Coimbatore.Interested in the research area of cloud Computing and Machine learning.

**Mownika M** Studying B.Tech Information Technology in Coimbatore Institute of Technology, Coimbatore. Doing research in Cloud computing and Cryptography.