PRIVACY PRESERVING OF MEDICAL DATA IN CLOUD USING PAILLIER ENCRYPTION TECHNIQUE

[1]P. Maragathavalli, [2]B. Praveen Kumar, [3]Shubham Kumar, [4]M. Sunitha
[1]Assistant Professor, [2]Final Year B.tech, [3]Final Year B.tech, [4]Final Year B.tech
[1,2,3,4] Information Technology
[1]Pondicherry Engineering College, Puducherry, India

Abstract: Cloud computing denotes an architectural reposition toward thin clients and conveniently centralized provision of computing resources. Client's lack of direct resource control in the cloud prompts concern about the potential for data privacy violations, especially leaking of sensitive data. Each cloud service will exchange data with other cloud, so when the data is exchanged between the clouds, there exist the problem of vulnerability of privacy. Hence the privacy problem about individual or company is inevitably exposed when releasing or sharing data in the cloud service. There are some privacy problems that are need to be addressed in the cloud computing. The first problem is the disclosure of sensitive private information when exchanging data through the cloud service. The second problem is that people getting inappropriate or unauthorized access to personal data in the cloud by taking advantage of lack of access control enforcement and security holes. Cryptography takes here a better place in preserving the privacy. In this paper, the privacy of user's data of personal health record is preserved using cloud computing which uses paillier to encrypt and decrypt the data. The individuals need not rely on their personal computer. The medical data documents can be easily and safely stored in the cloud itself. The proposed work shows that it is capable of giving better accuracy in comparison with other encryption mechanisms.

IndexTerms **– Privacy preserving, Data sharing, Cloud Service, Security issues, Sensitive data**

I. INTRODUCTION

Sensitive data is outsourced to the cloud there is a concern in regards with privacy. The Internet has evolved to a new phase. The traditional method of running the software on a desktop computer or server has become an old trend. Users are now moving towards the "Cloud" - an internet based computing with different services and variety of application in the mode of physical or virtual server because of internet growth. Cloud computing changed the way how an organizations managed their data because of its robustness, less cost and universally available nature. Medical data nowadays are stored in remote servers in cloud. Recent expansions of cloud computing gives us a heedful look at its real consequences involving confidentiality and privacy issues. During the outsourcing of sensitive data in the cloud, there is an agitation for the privacy of data that has been sent. By using encryption, the actual data cannot be obtained from the cloud server without proper access. Only the authorized users can get the data. Therefore cloud computing offers few incredible benefits for its users: the availability of enormous amount of software applications, faster processing power, unlimited storage, easy sharing and quick information processing [10]. The data can be easily accessed from anywhere with the help of internet by the end users. This might seem to be amazing but there remain threats for privacy, security, portability and reliability. The Cloud computing profound many benefits if the security along with privacy risks are captured and reduced. Many people nowadays started to use the power of the Cloud. The Cloud offers them so many facilities such as limitless flexibility in which users can access millions of software and databases if they have access, and they can merge them to customized services, users can easily find answers, share their knowledge, stream videos. Better security and reliability in which users can be free from worrying if their hard disk or pen drive is crashed. Enhanced collaboration in which the final user can share their knowledge and software via online, the Cloud provides users a ways to play and work together. Portability in which the ultimate user can access their data and tools wherever and whenever if they have an Internet connection. Simpler device facility in which the cloud eliminates the need for powerful tools and storage since we can store everything in cloud. They can interface with, a PDA, personal video recorder and cell phones. How the privacy of a user is preserved is shown in figure 1.1. During sharing of data via the cloud, the data owner shares the data to the requested user by encrypting the data. After encryption of data, the file is sent to the cloud. At the receiver end, the receiver searches for the file that is needed and the matched results are given back to the data user. The user has to decrypt the file using the key.

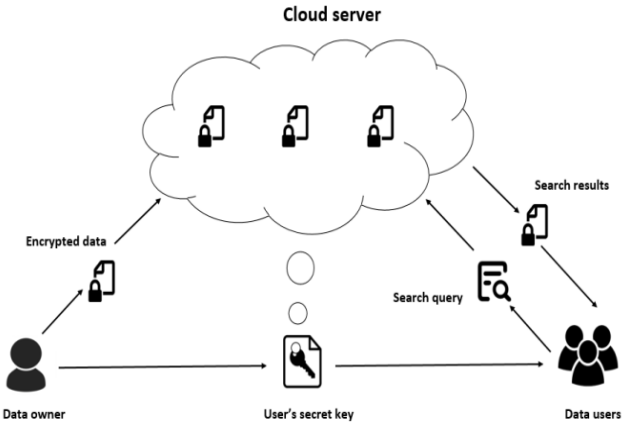

Fig -1.1: Privacy preserving in cloud environment

II. LITERATURE SURVEY

In this paper, the various methods have been proposed and many research is going on the area of enhancing the privacy and security in cloud computing. Some of the existing works are discussed in this paper.

2.1 Cipher text-Policy Attribute-Based Encryption with Delegated Equality Test in Cloud Computing

Qiang Wang [6] proposed that in cloud computing, cloud users can not only acquire useful data more effortlessly, but can offer noteworthy benefits to society as well by sharing their own data with other users or organizations. In this way, the cost for cloud users to share data can be saved significantly. They take personal health record (PHR) system for example. Patients in PHR system can measure and gather their sensitive PHR information by using medical sensors. To share their PHR data with the physicians in the hospital or other patients with similar symptoms, patients can upload their PHR data to a cloud server. Based on the collected PHR data from various patient features with similar symptoms, one can evaluate his/her own health status accurately. Moreover, the physicians can treat such kind of disease more precisely by analyzing the PHR data from a group of patients. When these data such as e-mails, personal health records, financial transactions, are accessed by illegal entities including the cloud server itself, the data owner may suffer incalculable economic and reputational losses. Therefore, every data owner should take measures to ensure the efficient access control of their data before uploading them to clouds. Attribute –based encryption (ABE) is commonly considered as a flexible and versatile solution to enforce access control with fine-granularity over encrypted data in the cloud computing [6]. So far, there are two types of ABE schemes, i.e., the ciphertext-policy ABE (CP-ABE) and Key-policy ABE (KP-ABE). In CP-ABE, any user is labeled with a set of attributes. One secret key can be used to decipher a specific ciphertext only if the attributes related to this secret key satisfy the policy embedded into the ciphertext. The limitation in this paper is that CP-ABE has limitation in terms of specifying policies and managing user attributes. The cipher faces data loss during the intrusion. Decryption keys only support user attributes that are organized logically as a single set, so the user can only make use of all possible combinations of attributes in a single set provided in their keys to assure policies.

2.2 Privacy-Preserving Deep Learning via Additively Homomorphic Encryption

Le Trieu Phong [1] proposed a model that a privacy-preserving deep learning system in which many learning elements performs neural network-based deep learning over a combined dataset of all, while not revealing the user's native information to a centralized server. Then the problem is solved by building an enhanced system with the following properties: 1) no information is leaked to the server and 2) accuracy is kept intact, compared with that of the ordinary deep learning system also over the combined dataset [1]. Their system merges machine learning and cryptography together. The system utilizes asynchronous stochastic gradient descent as applied to neural networks, along with additively homomorphic encryption technique. They show that their usage of encryption adds tolerable overhead to the ordinary deep learning system. Massive data collection, while vital for deep learning, raises the issue of privacy. Individually, a collected photo can be permanently kept on a company server, outside the owner's control. Legally, privacy and confidentiality concerns may prevent hospitals and research centers from sharing their medical datasets, barring them from enjoying the advantage of large-scale deep learning over joint datasets.

2.3 Privacy-preserving logistic regression with distributed data sources via homomorphic encryption

Y. Aono [8] proposed a system with dealing sensitive or private data, cares are necessary to preserve the privacy of the data. Here, a secure system for privacy protecting both training and protecting the data in the logistic regression via homomorphic encryption is proposed [8]. Regardless the non-polynomial tasks of training and predicting in logistic regression. We show that only additively homomorphic encryption is needed to build our system. Logistic regression is a standard method in supervised machine learning to classify data. It is widely applied in various fields of science and engineering. In several tasks using the

logistic regression, data contributors employ geographically distributed devices, raising the need of a central server to receive, store, share and process the data.

2.4 Input and Output Privacy-Preserving Linear Regression

Qiang Zhu [7] the storage and computation on the cloud can be seen as storing data and performing computations on a huge and globally available machine. Formally, a definition of cloud computing is given by NIST is saying that it is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources which can be quickly provisioned and released with stripped- down management effort or service provider interaction. Among others, machine learning is benefited from the development of cloud computing. Indeed, current commercial platforms such as the Amazon Machine Learning or the Google Cloud Machine Learning Platform allows clients to upload data to cloud servers to do various machine learning tasks [7]. The benefits of cloud computing come with threats, typically one of which is that plain data stored in the cloud may be accessed unwillingly. Promisingly, homomorphic encryption can balance the situation, as it enables input data secrecy and the computations over the data even in encrypted form.

2.5 Privacy-Preserving Public Auditing for Shared Data in the Cloud

Wang B [4] proposed that it is usual place for data to be not only stored in the cloud, but also shared among multiple users. Regrettably, the integrity of data in cloud is subject to skepticism due to the availability of hardware/software miscarriages and human errors. Various methods have been developed to give both public verifiers and data owners effectively audit cloud data integrity without retrieving the entire data from the cloud server. Public auditing on the integrity of data being shared with these existing mechanisms will unavoidably shows private information identity privacy to public verifiers [4]. It is standard for users to hold cloud storage services to share data with others in a group, as data sharing becomes a standard feature in most cloud storage offerings, including Dropbox, iCloud and Google Drive. The data integrity in cloud storage, however, is subject to doubtfulness and inspection, as data stored in the cloud can easily be lost or corrupted due to the unavoidable hardware/ software failures and human errors. To make this even worse, cloud service providers may be unwilling to inform users about these data errors in order to maintain the reputation of their services and avoid losing profits. Hence, the integrity of cloud data should be verified before any data utilization, such as search or computation over cloud data. The traditional approach for checking data correctness is to retrieve the entire cloud data, and then checks the integrity of data by verifying the correctness of signatures.
The limitation in this paper is that does not support traceability which makes it difficult to find the data.

2.6 Secure identity-based data sharing and profile matching for mobile healthcare social networks in cloud computing

Qinlong Huang [5] Mobile healthcare is an unconventional blend of mobile communication and mobile devices technologies, for it can provide necessary health information, routine care improvements, potential infectious disease prevention, health interventions, etc. It is getting more and more widely to apply the emerging cloud computing technology into the fields of mobile healthcare [5]. By using mobile healthcare system, the electronic health record (EHR) can be transferred over the network to the cloud service provider (CSP) for remote storage. Moreover, the healthcare providers can read it from an end device to provide real-time medical treatment. Mean-time, people tend to share and distribute the healthcare information via social networks, since social media is an extension of the healthcare professional and patient relationship. Consequently, mobile healthcare social networks (MHSN) are created for communicating patients so that they could share healthcare information using their mobile devices, and also connecting doctors and specialists for better healthcare. People in MHSN can interact with each other before making healthcare decision.

The limitation in this paper is that it requires a centralized server [11]. IBE's centralized approach implies that some keys must be created and held in escrow -- and are therefore at greater risk of disclosure.

III. OVERVIEW OF THE PROPOSED SYSTEM

3.1 Problem Definition

In the cipher text-Policy Attribute-Based Encryption with Delegated Equality Test has limitation it suffers specifying policies, managing user attributes and data loss occurs during encryption. With the help of Paillier Cryptosystem technique, this issue can be eliminated.

3.2 System Model

In our model, there are three entities are involved, as illustrated in Fig.3.2.1. They are data owners, the cloud server and data users. To Protect the Privacy of data in cloud, we are using a type of key pair based cryptography called paillier cryptosystem using which the user's data are being protected by encrypting them. Only those who have the corresponding private key can decrypt the ciphertext and will have access to data. In this way, the unauthorized users are eliminated from accessing the data. In order to check the accuracy of our system, linear regression is used. We are encrypting personal health record (PHR) which is sensitive data. Patients share their data over the cloud. While sending the information, the data is encrypted and the resulting ciphertext is generated and stored in cloud that it is not accessible to everyone. The encrypted data is

called cipher which is stored in the cloud. The person who has corresponding secret key of the message can only decrypts the data and can access it. The error rate is calculated using linear regression [9] and the accuracy of the system is being calculated which helps in finding the robustness of the system

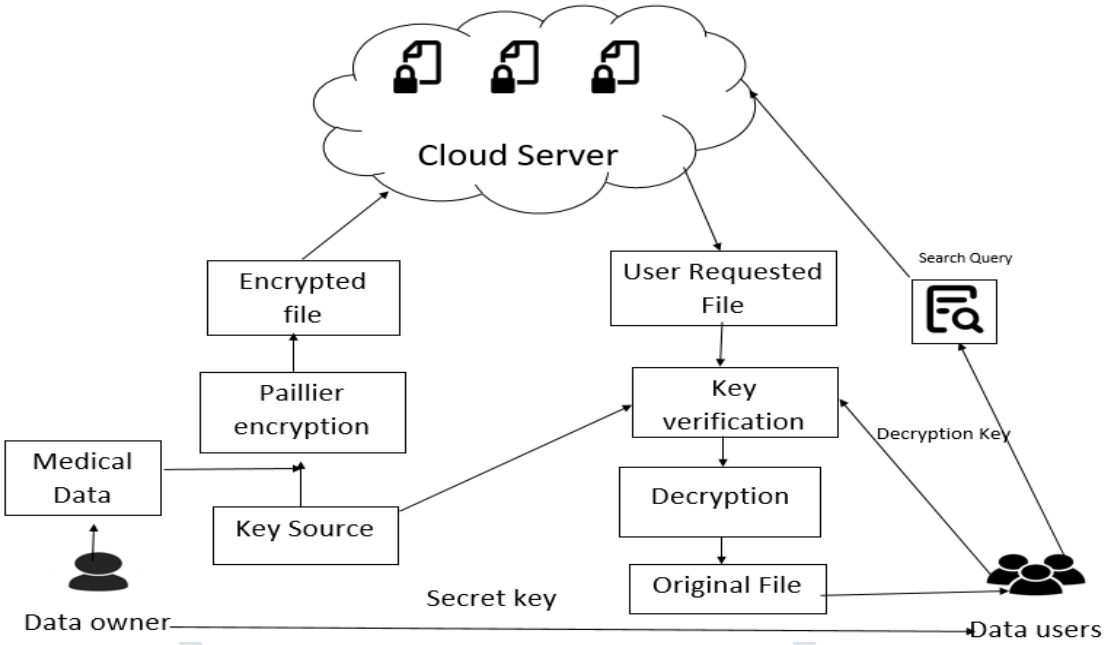

Fig - 2.1: Architecture Diagram

3.3 Paillier Encryption Technique

It is a type of key pair-based cryptography. Every user will have a private key and public key and messages encrypted by a public key can be decrypted only with the associated private key. One unique thing about paillier is that they provide additive homomorphism which means during encryption messages can be added with them and they will decrypt correctly. There are three main steps in paillier encryption technique. They are key pair generation, encryption and decryption [3].

i. ***Key generation process***

1. Select two large prime numbers *p* and *q* randomly and independently of each other such that

$$\gcd = (pq, (p-1)(q-1)) = 1.$$

2. Calculate n=p*q and take lcm and name it **λ.**

$$lcm = ((p-1)(q-1))$$

3. Choose a random integer g where it belongs to Zn^2.
4. The order of g is a multiple of n in Z* n^2 [Z* n^2 being the units, or invertible elements, of Zn^2]..
5. The public (encryption) key is (n, g).
6. The private (decryption) key is (λ,μ)

ii. ***Encryption process***

1. Let m be a message to be encrypted where $0 <= m < n$.
2. Select random variable r where 0<r<n and r belongs to $\mathbb{Z} * n^2$.
3. Compute ciphertext as

$$c = gm * rn \bmod n^2.$$

iii. ***Decryption process***

1. Let c be the ciphertext to decrypt, where c belongs to $\mathbb{Z}^*$ n^2.

2. Compute the plaintext message as $= L(c^{\lambda} \bmod n^2)$

3.4 Input & Output

Input- Person-level clinical data on patients admitted for acute pancreatitis Oct 2005-Sep 2009 in Veterans Health Administration hospitals. Output –Retrieval of original data.The dataset is distributed for training and testing which is 80% and 20% respectively. The data distribution is given in Chart 1.

Fig - 3.4.1: Dataset Distribution

3.5 Modules Description

3.5.1 Train the model using plaintext

In this module first we are loading the dataset into the program. After loading the dataset we need to normalize the dataset so that the machine can understand the dataset. We then reshape the data according to our requirements. Finally we partition the data into training set and testing set.

3.5.2 Data Encryption Module

In this module the data is encrypted using paillier encryption technique. During encryption corresponding public key and private key are generated. After encryption they sent it to the cloud server.

3.5.3 Data Decryption Module

In this module the encrypted data is decrypted using public key and retrieved from cloud. After decryption the accuracy of the system is measured.

3.6 Complexity involved in the proposal
Securely generating and distributing the secret keys. Checking the correctness of data.

4. EXPERIMENTAL ANALYSIS

4.1 Simulation Environment

The implementation of our proposed system was carried out using the jupyter notebook and python 2.7 software running on a personal computer with a 2.07 GHz Intel (R) Core (TM) I3 CPU, 4 GB RAM and Windows 10 as the operating system.

4.2 Security Analysis

In this model, the health record was obtained from the Veterans Health Administration hospitals. The data's are encrypted and stored in the cloud. They are retrieved and decrypted in order to find the error rate and accuracy of our work. The figure 4.2.1 depicts the input data of the personal health record. Table 4.2.1 depicts the description of the attributes of the given input data. Figure 4.2.2 depicts the sample encrypted data and figure 4.2.3 depicts the sample decrypted data.

	PatID	AP	LOS	sex	age	female	mi0	chf0
0	110193	1	5	1	68	0	1	0
1	108362	1	2	1	69	0	0	0
2	107797	1	5	1	47	0	0	0
3	100920	1	14	1	66	0	1	0
4	112456	1	1	1	52	0	0	0
5	101225	1	2	1	81	0	0	0
6	104164	1	3	1	54	0	0	0
7	107989	1	3	1	66	0	0	0
8	105977	1	1	1	56	0	1	1
9	106956	1	4	1	59	0	0	0

Fig - 4.2.1: Sample Input data

Table -4.2.1: Attribute Description

ATTRIBUTE	DESCRIPTION
PatID	Sequence ID based on random sort order
AP	Acute Pancreatitis = yes (Always true)
LOS	Length of stay initial AP admission, truncated at 365 days
Sex	F,M
Age	Age in years truncated at 90
Female	1 = yes, 0 = no
Mi0	Myocardial infarct year 0 (pre-AP)
Chfo	Congestive heart failure year 0 (pre-AP)

	PatID	AP	LOS	sex	age	female	Mi0	Chf0
0	BdnZ_yjR9tQHqt_m0KTeBQ==	jKGCn0vc__P2hfcdu8irRQ==	WY7mUBQ2rhsTpzRv-p_GQ==	jKGCn0vc__P2hfcdu8irRQ==	7Q6eSkxVvvhmpPZ0LG2KFQ==	BMkOj3E1mzY=	20954	_i9VOr5J0N0rThV04jTRtw==
1	Zqr3-4eb75XXmbX2dBHBvw==	jKGCn0vc__P2hfcdu8irRQ==	TyD0ZBpVA6JTrx8xIf18fA==	jKGCn0vc__P2hfcdu8irRQ==	8Om1LqWbCVApwoCymON2vA==	EeePrPm63Y8=	18213	_i9VOr5J0N0rThV04jTRtw==
2	Gx-bUYX5WD7s7weTgvOUw==	jKGCn0vc__P2hfcdu8irRQ==	WY7mUBQ2rhsTpzRv-p_GQ==	jKGCn0vc__P2hfcdu8irRQ==	eVTEk2o6c8FTR1monq8o3w==	BW1fcfWuFvE=	18213	_i9VOr5J0N0rThV04jTRtw==
3	SyuMbnZAk4dWaIO4KOZ13w==	jKGCn0vc__P2hfcdu8irRQ==	ffF28E_oVKxEIwqSsQAuDg==	jKGCn0vc__P2hfcdu8irRQ==	htRMVZCTuiCbFsZzC5Sitw==	I/5Bi56ZxKk=	20954	_i9VOr5J0N0rThV04jTRtw==
4	X2nrSRVVIX4xmpkTjwOftg==	jKGCn0vc__P2hfcdu8irRQ==	jKGCn0vc__P2hfcdu8irRQ==	jKGCn0vc__P2hfcdu8irRQ==	pguDqudZ8nbL9wWD7Vp00w==	J5Qt/K7zFAQ=	18213	_i9VOr5J0N0rThV04jTRtw==
5	NmchLYN9JX9RTeDB1qtJwQ==	jKGCn0vc__P2hfcdu8irRQ==	TyD0ZBpVA6JTrx8xIf18fA==	jKGCn0vc__P2hfcdu8irRQ==	tJcrBr7frH0PNQwReJjBvg==	IsM1r2H2Skl=	18213	_i9VOr5J0N0rThV04jTRtw==
6	snEdX-w-crGKWiOuOdSZM8==	jKGCn0vc__P2hfcdu8irRQ==	GAEvxDq6-9GhGQGqw4iOg==	jKGCn0vc__P2hfcdu8irRQ==	qxIfLoDZ7p8G1FwMVHgEGA==	D7H5dXM36dE=	18213	_i9VOr5J0N0rThV04jTRtw==
7	RNIUYwPbz7I0X7XUKNKnTA==	jKGCn0vc__P2hfcdu8irRQ==	GAEvxDq6-9GhGQGqw4iOg==	jKGCn0vc__P2hfcdu8irRQ==	htRMVZCTuiCbFsZzC5Sitw==	EXindx/G8o5w=	18213	_i9VOr5J0N0rThV04jTRtw==
8	gpcIPDIReO8vZtRu9OKr7Q==	jKGCn0vc__P2hfcdu8irRQ==	jKGCn0vc__P2hfcdu8irRQ==	jKGCn0vc__P2hfcdu8irRQ==	mNCTM6fb8KfguWKVkBtNA==	BIgGQxmj6U=	20954	jKGCn0vc__P2hfcdu8irRQ==

Fig - 4.2.2: Sample Encrypted Data

	PatID	AP	LOS	sex	age	female	mi0	chf0
0	110193	1	5	1	68	0	1	0
1	108362	1	2	1	69	0	0	0
2	107797	1	5	1	47	0	0	0
3	100920	1	14	1	66	0	1	0
4	112456	1	1	1	52	0	0	0
5	101225	1	2	1	81	0	0	0
6	104164	1	3	1	54	0	0	0
7	107989	1	3	1	66	0	0	0
8	105977	1	1	1	56	0	1	1
9	106956	1	4	1	59	0	0	0

Fig - 4.2.5: Sample Decrypted Data

The below chart 2 depicts the time taken by the proposed model compared with the CP-ABE encryption technique.

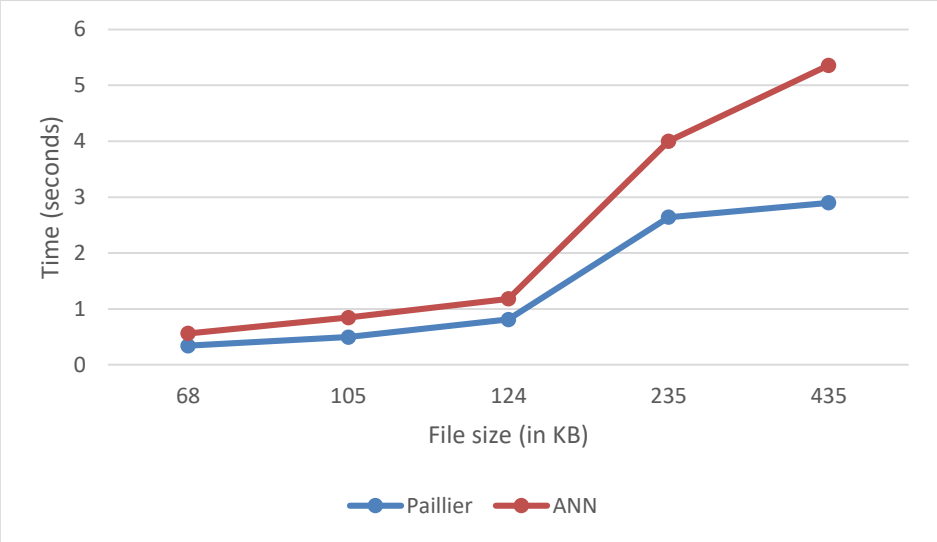

Fig - 4.2.6**:** Time taken by the model

The below chart 3 depicts the accuracy of the proposed model compared with the CP-ABE encryption technique.

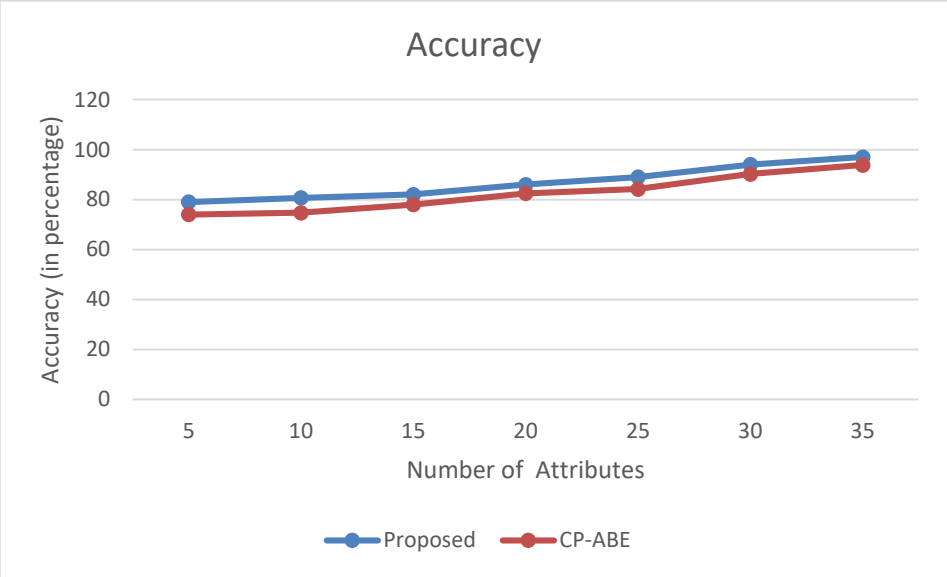

Fig - 4.2.7: Accuracy of the model

5. CONCLUSIONS

In our proposed work, the data is shared from the data owner to the data users via cloud efficiently achieving the data privacy while transmission of any confidential messages through the virtual server. This enables the users to share any confidential data through Paillier cryptosystem which is a complex crypto graphical structure enables the user to encrypt and decrypt the data without any vulnerability or data leakage. The accuracy rate is verified with the linear regression which gives an efficient accuracy rate.

In our future work, we are going to apply the same system of cryptography to images also. So that any confidential images that are need to be shared via the virtual servers can be securely encrypted using the paillier cryptosystem.

REFERENCES

[1] Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Lihua Wang, Shiho Moriai, "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption", IEEE Transactions on Information Forensics and Security, Volume 13, No. 5, pp. 1333–1345, May 2018.

[2] R.Nitya Lakshmi, R.Laavanya, M.Meenakshi, Dr.C.Suresh Gana Dhas, "Analysis of Attribute Based Encryption Schemes", International Journal of Computer Science and Engineering Communications, Volume 3, Issue 3, pp. 1076–1081, 2015.

[3] Payal V. Parmar, Shraddha B. Padhar, Shafika N. PatelNiyatee I. Bhatt, Rutvij H. Jhaveri, "Survey of Various Homomorphic Encryption algorithms and Schemes", International Journal of Computer Applications, Volume 91, No 8, pp. 26–32, April 2014.

[4]Wang B, B. Li and H. Li, "Oruta: privacy-preserving public auditing for shared data in the cloud", *IEEE Transactions on Cloud Computing*, vol. 2, no. 1, pp. 43-56, Jan.-March 2014.

[5] Qinlong Huang, Wei Yue, Yue He, Yixian Yang, "Secure identity-based data sharing and profile matching for mobile healthcare social networks in cloud computing," IEEE Journals and Magazines, Volume 6, pp. 36584–36594, July 2018.

[6] Qiang Wang, Li Peng, Hu Xiong, Jianfei Sun, Zhiguang Qin, "Ciphertext-Policy Attribute Based Encryption with Delegated Equality Test in Cloud Computing", IEEE Transactions on Information Forensics and Security, Volume 13, no. 5, pp. 1333–1345, May 2018.

[7] Qiang Zhu, Xixiang L, "Input and Output Privacy-Preserving Linear Regression", IEICE Transactions Information & System, Volume 100–D, No.10, 2017.

[8] Y. Aono, T. Hayashi, L. T. Phong, and L. Wang, "Privacy-preserving logistic regression with distributed data sources via homomorphic encryption", IEICE Trans. Inf. Syst., vol. 99-D, no. 8, pp. 2079–2089, 2016.

[9] Yongsen Han, Zhonghua Li, Huan Zheng, Wenmin Guo "A Decomposition Method for the Total Leakage Current of MOA Based on Multiple Linear Regression", IEEE Transactions on Power Delivery, pp. 1422 - 1428, 2015.

[10] P. Sreekumari, "Privacy-Preserving Keyword Search Schemes over Encrypted Cloud Data: An Extensive Analysis", *2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS)*, pp. 114-120, May2018.

[11]Girish, Phaneendra H.D, "Input Identity-Based Cryptography and Comparison with traditional Public key Encryption: A Survey", International Journal of Computer Science and Information Technologies, Volume 5, pp. 5521-5525, 2014.