**IJCTA,** 10(03), 2017, pp. 91-100


# A Framework on Security and Privacy-Preserving for Storage of Health Information Using Big Data

**J.L. Joneston Dhas* S. Maria Celestin Vigila** and C. Ezhil Star*****

***Abstract :*** Big data is a large amount of information. It has high value, volume, velocity, veracity, variety and variability. In the digital world information are generated and collected at very high range. According to IBM and Cisco recent report 2.5 quintillion bytes of information are generated daily and it will go up to 40 yottabytes($10^{24}$) in the year 2020. The patient information from the hospital, information from the government, private organisation, banking, internet and mobile phones is stored as big data in the cloud. It has both sensitive and non sensitive information. Causing damage to the sensitive data will cause a serious damage to the people. As the size of the data increases, new challenges of security arise and it is difficult to provide security to the information. Traditional security mechanism is not suited for protecting the big data. So an advanced security method must be implemented to protect the information.
***Keywords :*** Big Data, Role Based Attribute Control, Attribute Based Encryption, Auditing.

## 1. INTRODUCTION

In olden days the patient information are maintained in the particular hospital as a paper record. All the records are updated in the hospital itself. Afterwards the patient information is maintained in the hospital server. Updating the patient information is done by the authorised person in the hospital itself. So challenges for the data are very less. But nowadays all the health data are stored as big data in the cloud [1]. Big data is a collection of structured and unstructured data. Since the size of the data increased day by day so big data will be the best choice of storing the data. It takes the information from the social network, sensor, organization etc. It has huge amount of information that will generated in every second. Every day it generates one pegga bytes of information. From this huge amount of information a simple, reduced and useful information is taken and it will be used for improve the business and to take the future decision. It reduces the cost of storing and process the data. It has very large data sets so it will be easy for the user to make decision. In the EHR (Electronic Health Record) the information is taken from the hospital and all the patient history. It will have all the clinical history about the patient from birth to the current time.

Millions of people health information is stored as a big data in the cloud. So the people need not to carry all the records with them and maintain individually and all the information is stored online. Since the records are maintained in a public cloud and transmitted through public line privacy is an important factor for the big data. Hiding the sensitive information from the unauthorized user is the main task of privacy protection. Vishakha V. Kharche et. al. [2] proposed a privacy protection for big data. The protection

* Research Scholar Department of Computer Science and Engineering Noorul Islam University, Kumaracoil 629180, Tamilnadu, India. *E-mail : joneston.jl@gmail.com*

** Associate Professor Department of Information Technology, Noorul Islam University, Kumaracoil 629180, Tamilnadu, India. *E-mail: celesleon@yahoo.com*

*** Assistant Professor, Department of Computer Science and Engineering, Arunachala College of Engineering for Women, Manavilai, Tamilnadu, India. *E-mail: ezhilstar@gmail.com*

includes protecting capture, sharing, find, information privacy, analysis, transfer and visualization. Processing and storing of data is not an easy task. Processing include encrypt, encode or conversion of data [3]. The basic task and types of storage information is shown in the figure1.

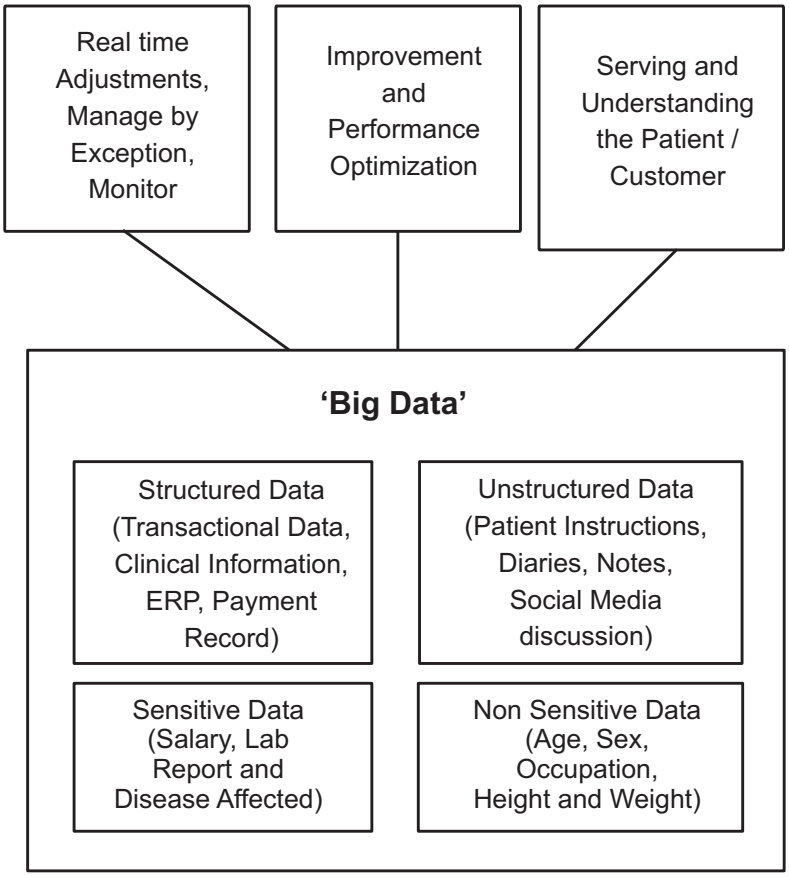


**Figure 1: Basic task done in big data which is apply in health data**

Processing, storing and retrieving the big data are a challenging task [4]. As the health information is sensitive information it should be more secure. But all type of people use the cloud to store and retrieve the data. Some of them are trusted people and some of them are unauthorised people. Unauthorised people may damage or modify the data [5]. Although all the data are encrypted if any modification will be done in the data the original data will also modified or they may inserted any malicious software to the data so the data will be damaged. Because of the damaged data it will not affect a particular person but it will affect millions of people. The data must be audited then and there. So the integrity of the data will be verified.

## A. Electronic Health Record

Electronic health record describes the patient information and treatment. EHR data centre collect large amount of patient information such as medication, clinical data, laboratory tests etc. EHR will be mined for clinical research for further treatment 6]. HIPAA (Health Insurance Portability and Accountability Act) gives control to the patient's medical record. It protects the privacy of the patient information. It explains how the medical records will be provided to the third parties and put penalties for those who not following the rules. As privacy view, attribute classified into three types

**Identity attributes :** Includes an individual which includes phone number, address, credit card number and name.

**Sensitive attribute :** Contains sensitive information like salary, lab report and disease affected.

**Non sensitive attribute :** Contains non sensitive information like age, sex, occupation, height and weight.

Before transmit the data to the third party the identity attribute will be removed called de-identification which protect the privacy of data. Sweeney [7] proposed 87% of United States population can be identified using the non-sensitive attributes like 5-digit zip code, date of birth and gender. Many existent platforms are used for sharing biomedical data. Currently there is much architecture used for biomedical data sets. Quantitative Imaginary Network (QIN) has The Cancer Imaging Archive (TCIA) which is designed for the development of candidate biomarkers and quantitative imaging methods for the measurement of tumour in clinical trial settings [8]. It shares the clinical data to multiple sites.

Global Alzheimer Disease Interactive Network (GAINN) communicates the big data to thousand of peoples in the world. The integrating Data for Analysis, Anonymization, and Sharing (iDASH) is a platform used by cloud to develop and share the algorithm and share the data in HIPAA [9]. tranSMART allows the expert users to communicate globally, they properly utilize the analytical tools which is best, promote the new transactional science in academic, pharmaceutical and non profit sector [10]. The Global Alzheimer's disease Interactive Network (GAAIN) creates a federated approach which will link the data to hundreds of thousands of participants who involve in research from all around the world.

### B. The Real Problems

There are many real time problems when we store the health record as a big data. The first is how a user will protect the information in the cloud. The next one is how to identify the record and how to protect the health information from the unauthorised user. The size of the data is the main challenge for big data. A large amount of Structured and unstructured data must be processed efficiently. The data are not in order and has one peggabyte of information are generated daily. Before five years itself US health data reached 150 exabytes($10^{18}$)[11]. For the large population countries like India and china already reached yottabytes($10^{24}$). The data is continuously generated from multiple sources and the data increased gradually. It provides metabolomics, proteomics, genomics, and clinical information about the individuals. Health data have meaningful datasets and it is too fast, big and complex to process.

Other challenges faced by the health information are speed, variety and heterogeneity of data. The system must mine, process the data and change to make decision making from that data. The data are heterogeneous from variable sources. The heterogeneous data are qualitative (demographics, text) and quantitative (laboratory tests, images, sensor data) and it should able to answer for clinical questions. The data is coming from different sources and there are different types of people use the cloud. Some are trusted and some are untrusted. Some people may spread malicious software to the data, damage the data, or modify the patient information and it will cause a serious damage to the patient. Cloud providers and existing data processing are not detected and monitor the data leakage. So privacy preservation, data auditing and data protection should be achieved for Electronic health information. So a public auditing should be done periodically and the integrity of the data is verified. An efficient access control mechanism should be provided to control the unauthorised user.

## 2. RELATED WORK

To secure the health data different techniques such as authentication, digital ware marking and MPEG encryption schemes are used. Some works are done in context aware policies and authorization and it is difficult for big data. Arthur W Toga1 and Ivo D Dinov [12] proposed an efficient method to access the big medical data. The set of agreement will be done between the owner and the user. So the data will be secure and the security metrics is high.J. B. Joshi et. al. [13] the multimedia data have various classifications depending in quality and/or nature of information. The authors provide a secured model for create and store the multimedia data. Bhatti et.al [14] proposed a problem for access management of multimedia data and propose a distributed access management for multimedia data.

E. Bertino et. al. [15] proposed GEO-RBAC (Spatial Characteristics with Role Based Access Control), defines spatial role, each spatial location in the organisation have their own role. When the number of roles

increases there is number of location in the organisation and it is an extension of RBAC (Role Based Access Control). J. B. Josh et.al [16] propose RBAC contains four components which includes set of roles, set of devices/users, set of sessions and set of permissions. A user may be an autonomous agent, a human being, a task, a subsystem or a physical device. In this method the access is based on the role of the user and the subject binds within the organization. It is also known as non-discretionary access control because the user gets the privileges given to his role. The user has no control over the assigned role. According to GST-RBAC, a role is enabled at a particular time and location but not at other times and locations. It takes the parameter time to validate the access request. It allow the specification of temporal and spatial constraints on user-role assignment, role-permission assignment, runtime events and activation GST-RBAC (Generalized Spatio Temporal Role Based Access Control Model) model with spatial and temporal constraints is available in [17].

Attribute-based encryption (ABE) [18], [19], [20] provides the security for the end to end data in the cloud. The data owner should define the access policies and decryption is done under that policy. So the user who is having the above attribute can only decrypt. More number of users use the data from the cloud. So the access policies are changed frequently. The policy updating is not available in attribute based access control. It is one of the suitable access controls for big data in cloud. There are two types of attribute based encryption, key policy attribute based encryption(KP-ABE) [18] and cipher text policy attribute based encryption (CP-ABE) [19] [20]. In the case of CP-ABE a particular set of data will be encrypted. CP-ABE describes the access policy and user attribute are attached in the encrypted data. KP-ABE describes the access policies and encrypted data of the attributes are built in user secret key.

Attribute based access control (ABAC) [21] [22] provides the data confidentiality for big data in the cloud. In this the data owner defines the access structure and encryption is done based on the access structure. The data owner defines the attributes so the users possess to decrypt. When using the ABE to develop the access control the policy updating is difficult. Since the data are stored into the cloud and not in the local system. V. Goyal et.al [18] Key-Policy Attribute Based Encryption (KP-ABE). In this the cipher text has a set of private keys and attributes and it associated with access structures. In this a selected cipher texts can be decrypted by the user. Changji Wang et. al.[23] Propose a Key-Policy Attribute-Based Encryption Scheme with Constant Cipher text Length. In this they constructed a new KP-ABE scheme which supports any monotonic access structure with constant size cipher text. But the private key grows as multiple sizes in the desired ABE systems. The comparison of the different access control is shown in the table 1.

**Table 1**
**Comparison of different access model**

| *Access Control* | *Policy Updation* | *Policy Updation without reencrypt* | *Spatial Extention* |
|---|---|---|---|
| GEO-RBAC | Yes | Yes | Yes |
| RBAC | Yes | No | No |
| ABAC | No | No | No |

Access policy updation is done for secure access of big data. The cipher text is frequently updated to the new access policy and proposed in [24][25]. Kan Yang et. al. [24] propose an access policy updation method without reencrypt the ciphertext. So the computation of big data is highly reduced. Also it performs its work in the cloud server itself. So the communication overhead is avoided between sever and the user. Privacy preserving is an important factor for the health record. All the users are not disclosing their sensitive information like disease, age, salary etc. to others and they want to keep it secret. The original information is stored as a private table which contains multiple records. Each type of record contains four attributes.

**Identifier :** Identify the person directly and contains the information such as mobile number, ID and name.

**Quasi Identifier :** Linked with the external table and the individual records are re-identified.

**Sensitive Attribute :** Contains the sensitive information such as salary, disease, age etc. And the information need to be protected from the third party.

**Non Sensitive Attribute :** Contains the attribute which is other than sensitive attribute, quasi identifier and identifier.

Before releasing the information the data is anonymized, this means removing the identifier and modifying the quasi identifier. So the original information are hidden and the adversaries cannot able to identify the information. Rongxing Lu et. al. [26] proposes the privacy preserving mechanism in big data. Also they propose the privacy requirements for data collection, storage and processing. Fung et. al. [27] proposed several privacy models and also several anonymization.

**Generalization :** Replace the some of the original value.

**Suppression :** Replace the original value with special symbol.

**Anatomization :** It will not modify the value rather it modifies the relationship of two attributes.

**Permutation :** It partitioning the data into several groups and then shuffling the values.

**Perturbation :** The original value of the data is replaced with synthetic values.

**Table 2**

**Original table of the Patient**

| *Age* | *Sex* | *Zipcode* | *Disease* |
|---|---|---|---|
| 5 | Male | 600013 | HIV |
| 12 | Female | 629805 | FLU |
| 8 | Female | 600890 | Gastrisis |
| 19 | Male | 340543 | Cancer |
| 16 | Female | 435654 | FLU |
| 14 | Male | 656809 | Cancer |
| 15 | Female | 600435 | Gastrisis |
| 11 | Female | 549654 | Cancer |
| 2 | Male | 234653 | FLU |

Table 2 shows the original value of the patient record. Table 3 shows the privacy preserving modified table so the third party cannot be able to identify the original patient record. Several anonymization techniques are implemented to protect the privacy information. Qingchen Zhang et. al. [28] proposed the privacy preserving done by deep computational model. In this case the computation is performed in the cloud server instead of the client system. So the computation overhead is reduced in the client system. Back propagation algorithm is used for privacy preserving so it process the data upto maximum iterations and the data is protected from the adversaries. Maria Celestin Vigila and Muneeswaran [29] implement a Elliptic curve to secure the sensitive information from the brute force attack and provides an efficient approach to protect the sensitive information from the unauthorized persons. Jemal H. Abawajy et.al [30] proposes a large iterative multitier ensemble which will combine multiple multitier ensemble classifiers and classify multiple informations in a simple manner which is used to identify the malicious software in the big data in a simple and efficient manner.

**Table 3**
**Privacy preserving table**

| *Age* | *Sex* | *Zipcode* | *Disease* |
|---|---|---|---|
| [1,10] | People | 6***** | HIV |
| [2,20] | People | 6***** | FLU |
| [1,10] | People | 6***** | Gastrisis |
| [2,20] | People | 3***** | Cancer |
| [2,20] | People | 4***** | FLU |
| [2,20] | People | 6***** | Cancer |
| [2,20] | People | 6***** | Gastrisis |
| [2,20] | People | 5***** | Cancer |
| [1,10] | People | 2***** | FLU |

## 3. PROPOSED SYSTEM

In the proposed system a secured system model is shown in the figure 2. It consists of the following entities: Owner, Users (Hospital), Authorities (AA), Data Receiver, Public Key ($P_U$), Private Key ($P_R$), Cloud Server, Access Control, Identify Receiver, Result Verification.

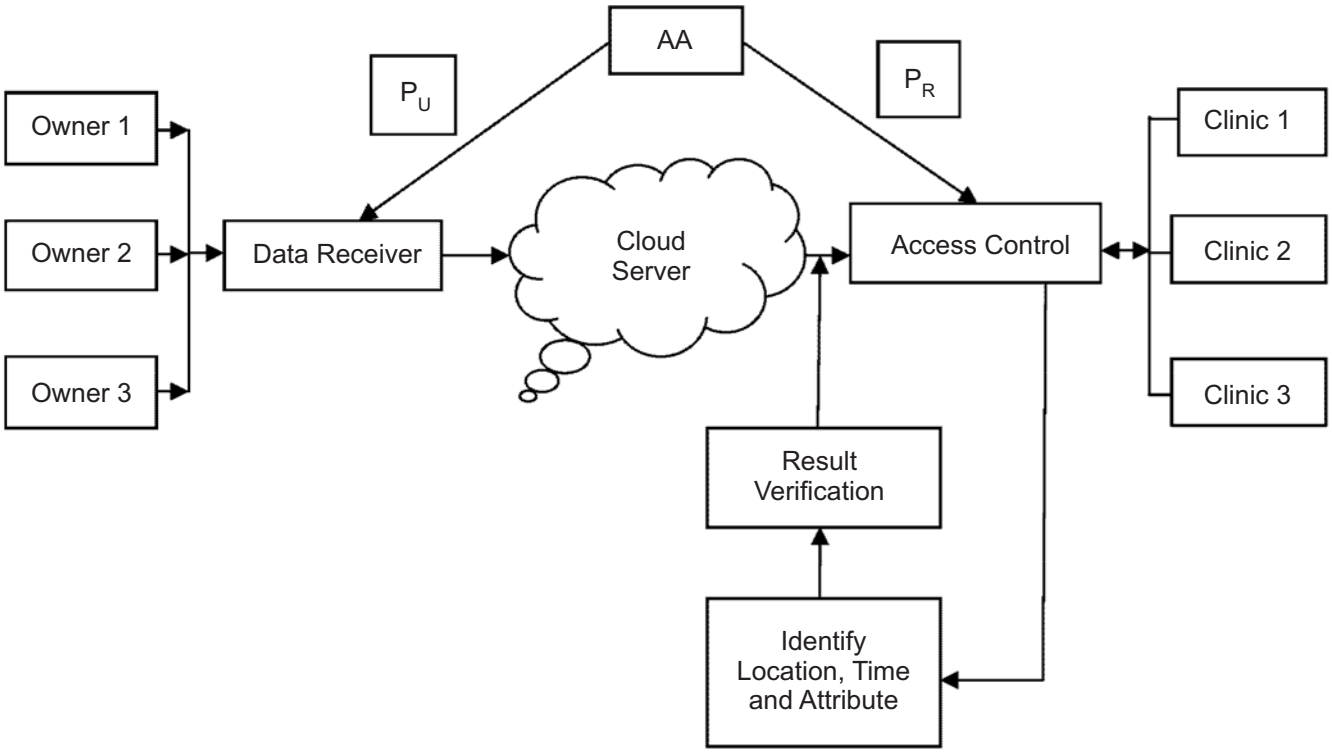


**Figure 2: Secured Cloud Server Framework for Health Record**

**Cloud Server :** The owner of the data stores their information in the cloud server and it performs various operations such as searching, auditing, privacy preserving, updating the cipher text policies etc.

**Owners :** People who will generates the data and upload all the information to the cloud server. A person can authorize or deny access. Under the policy he encrypts the data and stores it in the cloud server.

**User :** Each user is having an identity and he can take the cipher text from the cloud server and decrypt the cipher text only if all his attributes are satisfied with the access policy.

**Authority :** It has an authority id and generates a private key and public key and distribute to the user and the owner.

**Data Receiver :** It receives the data from the data owner for further pre-processing like normalization, transformation, cleaning, and feature extraction.

**Map Reduce :** It is a frame work and the tasks are performed parallel in case of large data which is used by cloud server for processing big data and have Map and Reduce functions. The function Map splits the big data as <key, value> and it is generated by aggregating the input <key, value> in Map phase. Reduce generate the output <key, value> depending upon the intermediate<key, value>. So the user can easily retrieve the information. A framework is created to process the big data. A large number of users are logged in the data centres and all of them are processed by Map Reduce.

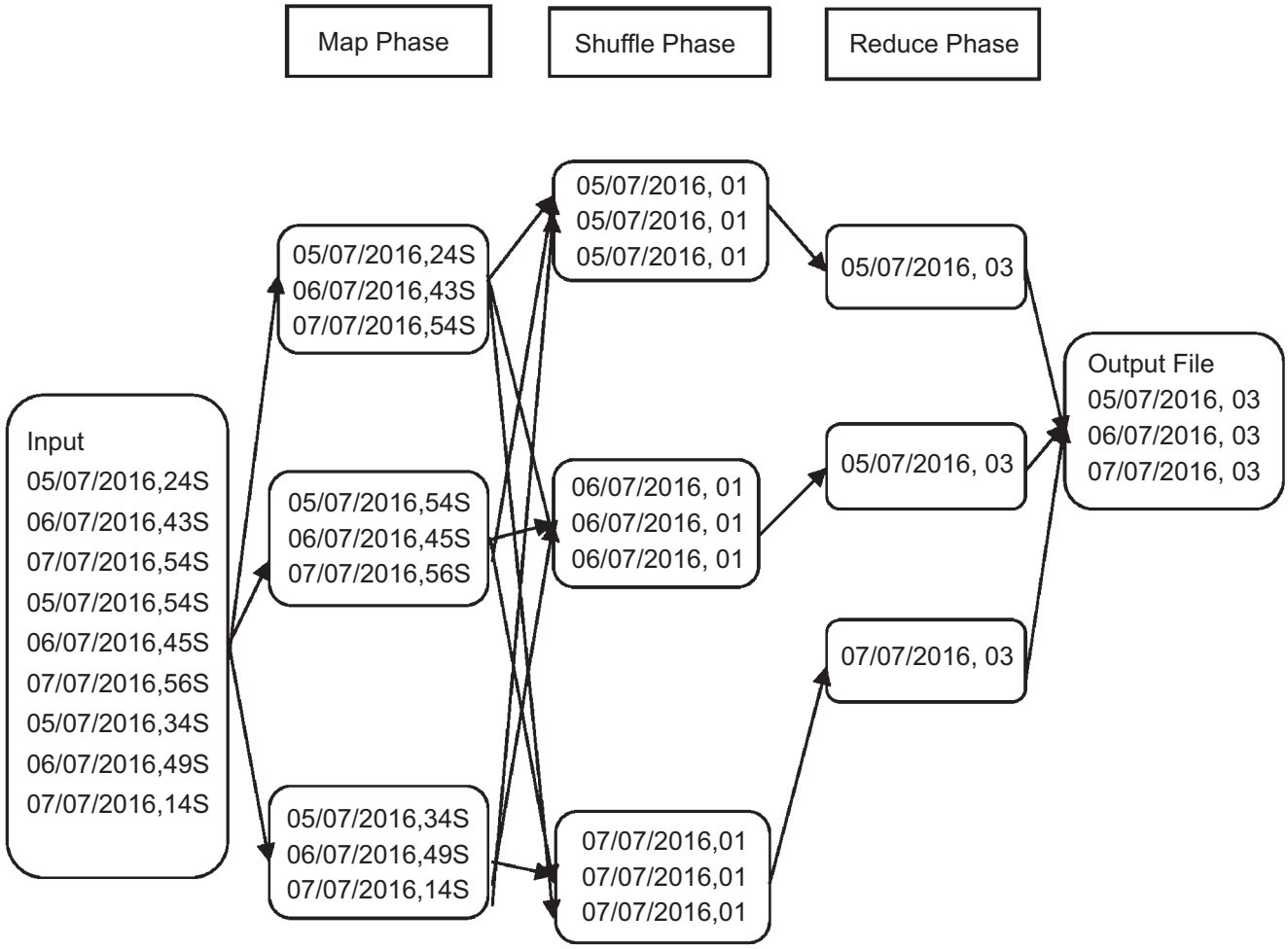


**Figure 3: Map Reduce**

The figure 3 explains the concept of Map Reduce. The size of the big data is large and variety so Map Reduce effectively handles the big data for processing and storing.

**Public Key :** It is the value provided by the certain authority and used for encrypting the data.

**Private Key :** It is the value provided by the certain authority and used for decrypting the data.

**Identity Receiver :** It receives the identity from the user to verify all his attributes, time and location.

**Result Verification :** It verifies all the attributes of the user. If it satisfies the user can able to decrypt the data.

**Access Control :** It identifies the authorized user from their attributes, time and location. If all the requirements are satisfied then only it provides the private key to the user to decrypt the message.

In this framework the owner will be a patient or the hospital where the patient takes the treatment. All the patient information will be taken regularly and it will be processed by the data receiver and it will do the several processes like encryption, compression, analysis etc. Authorization agent generates a public key and private key and distribute to data receiver and access control. After processing the data it will be stored in the cloud server. Data retrieval will be done in the clinic. It will have a strong access control and only the authorized person will be able to access the information. It will have a location attribute and the user will be able to access the information only in the particular time and location. So the unauthorized person will be able to access the information.

Once the person want to access the information first he will be work in a particular location and it will be identified through GPS (Global Positioning System) and some other object will also taken as the attributes. All the location and the time will be identified by result verification. Once all the attributes are verified as correct the decryption key will be given to the user to decrypt the needed information by the data receiver. This framework ensures an efficient process of big data and provides a good relationship between the patient and the doctor. Since all the patient information are available up to date. So the doctor can easily predict the patient disease and provide the treatment quickly and efficiently. It this framework all the patient records are securely stored and retrieved by an authorized persons only.

### A. Data Set

When the size of the data is smaller the data set STL-10, PeMS (Performance Measurement System) and DLeMP (Dalian Economy Management Platform) are affordable. Performance will be evaluated by PeMS and DLeMP. Traffic prediction is done by PeMS data set. Data are collected continuously from 8,000 locations and over 30,000 detectors. DLeMP is used for economy prediction and information is collected from 250 towns. It contains 500,000 items and has 52 attributes. The prediction will be accurate in PeMS. When the size of the data increases the training time will also increase. The training time includes encryption and decryption time and computation time. We can also implement parallel processing and increasing the cloud server to improve the performance. In PeMS three tasks are performed. *i.e*., it will predict the traffic in peak and non peak period. Predict the traffic of peak in several intervals. STL-10, PeMS and DLeMP are smaller than NUS-WIDE data set. When the number of nodes increases the performance will be improved and improves the capacity of the system. For the large data set NUS-WIDE is used and it improves its efficiency by adding more cloud servers. It supports more than 10 cloud nodes.

## 4. CONCLUSION

Big data is used to study the clinical research in the real world. It will provide a precision and effective medicine for patient stratification. This is the key task for personalized healthcare. It provides a better health services to the people. The improvement of big data in the area of bioinformatics, health informatics and sensing provides a better impact for clinical research. In the case of large population it needs large number of distributed inference and on-node data abstraction. From the past disease can forecast and prevent the same incident in future to others. All the data must be securely stored and it should be free from malware, unauthorized user, Access control policies must be updated for every time period for secured access. Data is audited and ensure the data integrity.

Big data is not used for only one field. It is used in all fields to improve their business. Predicting the future by using the current trend can be possible by big data. In the medical industry they find a lot of new diseases attack the people due to environment changes. So all the medical data will be stored and the medicine for such disease will be finds out. When any environment changes occur it will be compare by the old value if it assumes to be same then precausive measures will be taken to avoid those diseases. So the peoples will be secure from the diseases. All the big data information will be secure and audited regularly. In the case of GEO-RBAC model new attributes like uniform, any objects, etc. can also be included for controlling the access. Identification of malicious software can be done by five tier multiple classifier. So that all the sensitive information will be highly secure.

## 5. REFERENCES


1. Linquan Zhang, Chuan Wu, Zongpeng Li, Chuanxiong Guo, Minghua Chen and Francis C.M. Lau, "Moving Big Data to The Cloud", Proceedings IEEE INFOCOM, 2013.
2. Vishakha V. Kharche, Prof. Alokkumar Shukla, "A Security and Privacy Preserving in Big Data", IORD Journal of Science & Technology E-ISSN: 2348-0831 Vol. 2, Issue 3, pp. 32-37, 2015.
3. Hao Zhang, Gang Chen, Beng Chin Ooi, Kian-Lee Tan, and Meihui Zhang" In-Memory Big Data Management and Processing: A Survey", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 7, July 2015.
4. Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald and Edin Muharemagic, "Deep learning applications and challenges in big data analytics", Journal of Big Data, 2015.
5. R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era", IEEE Network, Vol. 28, No. 4, pp. 46–50, 2014.
6. C. Friedman, L. Shagina, Y. Lussier, and G. Hripcsak, "Automated encoding of clinical documents based on natural language processing", J.Amer. Med. Informat. Assoc., Vol. 11, pp.392–402, 2004.
7. Sweeney, L. "k-Anonymity: A Model for Protecting Privacy", International Journal on Uncertainty, Fuzziness and Knowledgebased Systems (10:5), pp. 557-570, 2002.
8. Kalpathy-Cramer J, Freymann JB, Kirby JS, Kinahan PE, Prior FW "Quantitative imaging network: data sharing and competitive AlgorithmValidation leveraging the cancer imaging archive", Transl Oncol 7(1), pp. 147–152, 2014.
9. Ohno-Machado L, Bafna V, Boxwala AA, Chapman BE, Chapman WW, Chaudhuri K, Day ME, Farcas C, Heintzman ND, Jiang X, Kim H, Kim J, Matheny ME, Resnic FS, Vinterbo SA, "iDASH: integrating data for analysis, anonymization, and sharing", J Am Med Inform Assoc 19(2), pp. 196–201, 2011.
10. Athey BD, Braxenthaler M, Haas M, Guo Y,"tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research", AMIA Summits Transl Sci Proc 2013 pp. 6–8, 2013.
11. M. Cottle,W. Hoover, S. Kanwal, M. Kohn, T. Strome, and N.W. Treister, "Transforming Health Care Through Big Data, Institute for Health Technology Transformation", Washington DC, USA, 2013.
12. Arthur W Toga1 and Ivo D Dinov, "Sharing big biomedical data", Toga and Dinov Journal of Big Data, 2015.
13. J. B. Joshi, Z. K. Li, H. Fahmi, B. Shafiq, and A. Ghafoor, "A model for secure multimedia document database system in a distributed environment", IEEE Trans. Multimedia, Vol. 4, No. 2, pp. 215–234, Jun.2002.
14. R. Bhatti, B. Shafiq, M. Shehab, and A. Ghafoor, "Distributed access management in multimedia IDCs Computer", No. 9, pp. 60–69, 2005.
15. E. Bertino, B. Catania, M. L. Damiani, and P. Perlasca, "Geo-RBAC:A spatially aware RBAC," in Proc. 10th ACM Symp. Access Control models Technology, pp. 29–37, 2005.
16. J. B. Joshi, E. Bertino, U. Latif, and A. Ghafoor, "A generalized temporal role-based access control model", IEEE Trans. Knowl.Data Eng., Vol. 17, No. 1, pp. 4–23, Jan. 2005.
17. A. Samuel, A. Ghafoor, and E. Bertino, "A framework for specification and verification of generalized spatio-temporal role based access control model", Center for Education and Research in Information Assurance and Security, Purdue Univ., West Lafayette, IN, USA, Tech. Rep. CERIAS TR 2007-08, 2007.
18. V. Goyal, O. Pandey, A. Sahai, and B. Waters, Attribute-based encryption for fine-grained access control of encrypted data," in Process 13th ACM Conference in Computation and Communication, pp. 89–98, 2006.
19. J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in Process IEEE Symposium Security Privacy, pp. 321–334, 2007.
20. B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization", in Proc.14th International Conference Practical Theory Public Key Cryptography Conf. Public Key Cryptography, pp. 53–70, 2011.
21. K. Yang and X. Jia, "Attributed-based access control for multiauthority systems in cloud storage", in Process IEEE 32nd International Conference in Distributed Computing System, pp. 1–10, 2012.
22. T. Jung, X.-Y. Li, Z. Wan, and M. Wan "Privacy preserving cloud data access with multi-authorities," in Process Conference in Information and Communication, pp. 2625–2633, 2013.

23. Changji Wang and Jianfa Luo, “An Efficient Key-Policy Attribute-Based Encryption Scheme with Constant Ciphertext Length”, Hindawi Publishing Corporation Mathematical Problems in Engineering Volume, 7 pages, 2013.

24. Kan Yang, Xiaohua Jia, and Kui Ren,” Secure and Verifiable Policy Update Outsourcing for Big Data Access Control in the Cloud”, IEEE Transactions on parallel and distributed systems, Vol. 26, No. 12, December 2015.

25. Kan Yang, Xiaohua Jia, Kui Ren, Ruitao Xie and Liusheng Huang,” Enabling Efficient Access Control with Dynamic Policy Updating for Big Data in the Cloud”, IEEE Conference on Computer Communications, IEEE Infocom 2014.

26. Rongxing Lu, Hui Zhu, Ximeng Liu, Joseph K. Liu, and Jun Shao,” Toward Efficient and Privacy-Preserving Computing in Big Data Era”, IEEE Communications Society, Vol. 28, pp. 46-50, 2014.

27. B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, ``Privacy-preserving data publishing: A survey of recent developments,'' ACM Computing Survey, Vol. 42, No. 4, June 2010.

28. Qingchen Zhang, Laurence T. Yang, and Zhikui Chen, “Privacy Preserving Deep Computation Model on Cloud for Big Data Feature Learning”, IEEE Transactions on Computers, Vol. 65, No. 5, May 2016.

29. S. Maria Celestin Vigila, K. Muneeswaran, “A new elliptic curve cryptosystem for securing sensitive data applications”, International Journal of Electronic Security and Digital Forensics, Vol. 5, No. 1, 2013.

30. Jemal H. Abawajy, Andrei Kelarev and Morshed Chowdhury, “Large Iterative Multitier Ensemble Classifiers for Security of Big Data”, IEEE Tranactions on Emerging Topics in Computing, Vol. 2, No. 3, September 2014.