Top-Down Approach in Access Control with Timing Enabled Key Distribution for Hierarchical Systems in Electronic Health Records

C. Kalpana

Research Scholar, Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology, Chennai,
Tamil Nadu, India
cm.kalpana@gmail.com

Dr. S. Revathy

Associate Professor, Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology, Chennai,
Tamil Nadu, India
Ramesh.revathy@gmail.com

Abstract

Secured transmission for accessing a large amount of data in today's technology is not very easy. To overcome this disadvantage, we are providing hierarchical access control with suitable algorithms. Sharing of health record data to remote places through the network has a lot of security and privacy issues. The proposed system is for sharing of datasets with the dynamic key distribution for electronic health records. Using supervised algorithms, we are creating a hierarchical system of datasets for the sharing of datasets in a tree structure to the user. The tree structure of datasets having timing enabled the key to share for limited-time access to overcome security issues. On the user side, using a supervised algorithm the users also converted into a hierarchical system based on their grade of access, the access is controlled by a timing enabled key distribution to the hierarchy of users. The level of access is ranking based which divides the datasets into different categories in a hierarchical pattern.

Keywords: access control; algorithm; attributes; classification; distributed key; ranking.

1. Introduction

Electronic Health Records (EHR) is used to replace the existing records in papers. The Electronic Health Record owners can search their files from anywhere from an anonymous system. They use online and offline to process data at a speed in which data is greater than previous systems to avoid time complexity. EHR Data having sensitive information related to individual patients. Access to such sensitive data by health providers, friends, and families in unusual situations will create security issues. Moving of health data from one database to another may lose the privacy of the individuals and also security problems may also arise. To overcome the electronic records access issues many types of access policies with access controls given in the existing systems. [5]They proposed an extensible access control markup language, using a cloud-based model with attribute-based access control. The proposed system is designed to use electronic signatures for the document user. The attribute values taken for access not defined using statistical methods. Time constraint is not used for access. The latest languages overcome this XML based model. [20]They proposed a system that is working offline and online to work at the highest speed than the existing system with lesser time complexity. It's a model for the hierarchical system but within a geographical region. [2, 19] This paper proposed distributed keys for the access of electronic health records based on cipher text policy with attribute encryption. [2, 9] This paper proposed a feature-based ranking for the access of medical data using random forest classifier to give the predictor.The model designed for the classification of datasets only but not user side access hierarchy. [11]A two-step combination of public-key encryption and DUKPT (Derived Unique Key Per Transaction) is used. Not patient-centric, it may affect privacy policies defined by patients. [21, 4] Papers proposed a model by giving role-based access to electronic health records. But they failed to give controlled access based on the ranking system to the users who are accessing from different environments and distance at different time limits. [3, 16] Based on analysis of different existing system they proposed an access control model for electronic health records for hierarchical key access system, it is the survey of different existing system for this paper proposed system. [17]The limited time based access to the patient data for the user to overcome the issue in patient stay for a particular period.

2. Proposed Work

Sharing of electronic health record data rises privacy problems to data owners to avoid this giving access control with time-bound access to the request is the proposed area of research. The access controls with spatial and temporal dimensions given for the hierarchical systems. This model aims in providing privacy since dynamic key creation and distributions are included based on the time constraint using a new key assignment scheme suitable for hierarchical systems. The access to the user from any sector will be allowed by the system with time-bound is useful to shift the patient from one place to another with privacy concerns.

3. Proposed Architecture

Figure 1 shows the different role based access to the electronic health records. The data requested is shared based on the type of access based on role. The based access is controlled by sharing the data based on attributes sensitivity. Different access made by the key generation to read or write on certain attributes of the data to overcome the security issues.

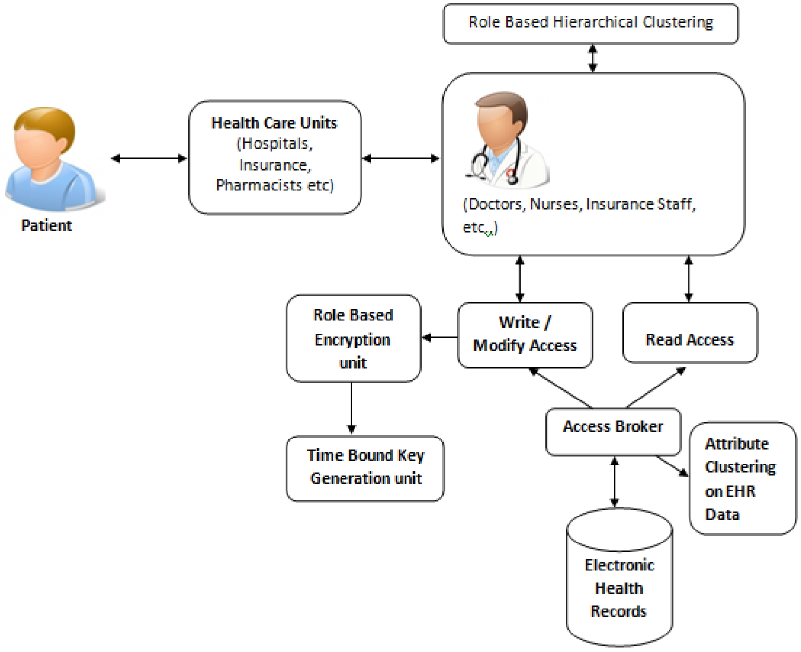

Role Based Hierarchical Clustering
Patient
Health Care Units (Hospitals, Insurance, Pharmacists etc)
(Doctors, Nurses, Insurance Staff, etc,,)
Role Based Encryption unit
Write / Modify Access
Read Access
Time Bound Key Generation unit
Access Broker
Attribute Clustering on EHR Data
Electronic Health Records

Figure 1: Access Control for Electronic Health Records

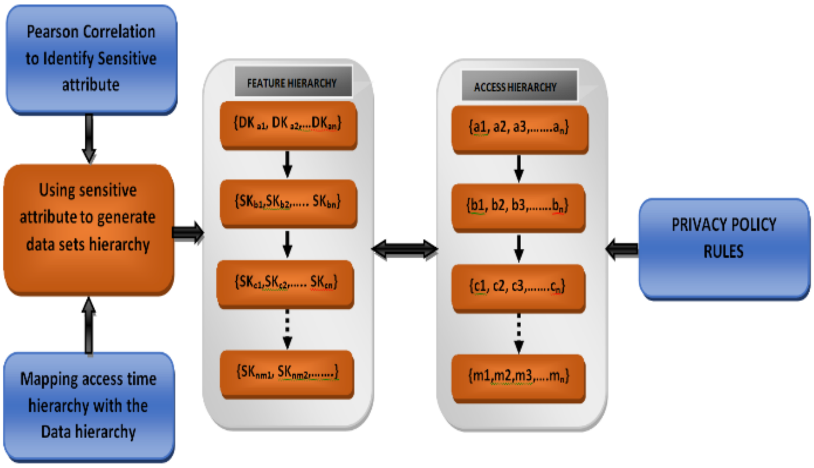


Figure 2: Hierarchical access control with timing enabled key distribution

Figure 2 is the architecture of the access control system for a hierarchical system with timing enabled distributed key control. Using Pearson correlation we can identify the sensitive attribute in the data set. The sensitive attributes made a hierarchy of access based on the ranking given to each attribute. A time hierarchy also mapped to the ranking access system. The feature selection method is used for feature hierarchy. In the access hierarchy based on the user's grade level, the access is controlled by using privacy policies with timing enabled public key distribution.

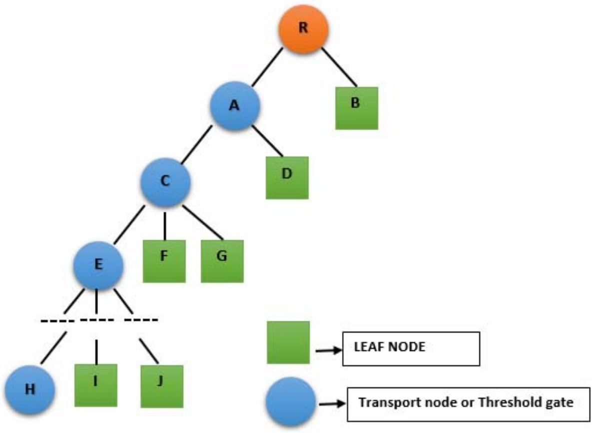


Figure 3: Tree structure with sensitive attribute-based access

The access authorization is set in a tree structure (T) which describes the user by using the attribute. The data user access is enabled only when the signature associated in tree matches with the user attribute. The access tree T is defined as (x, n(x), index(x), attribute(x), Th(x), parent(x)),
In Fig. 3, node A indicates the threshold gate. X represents a node of the tree T. Each node is a threshold gate, for example, "AND'" or "OR" gates, where attributes are associated by leaves.

- Parent(x)is the parent of node x. That is the parent (A) denoted by the root node R.
- n(x) is the number of children in the node xin tree T. In Figure 2, n(c) = 3.
- Th(x)is the transport node value forx, and 0 lessthan Th(x) less than or equal tox. For Th(x) is one, the transport node is taken as OR gate while Th(x)is equal to n(x)represents, it is an AND gate. If x is a leaf node, Th(x) = 1. Th(A) =2 means it is an AND gate.
- Attribute(x) represents it is an attribute value on the leaf node x in tree T.
- Index(x)represents the number integrated with x. Then the value is from 1 to n(x), in figure 2 the designated key is assigned to x for access.

4. Identification of Sensitive Attribute

Privacy threats related to three different types of attributes they are,the attributes that can identify individual's details, like address, patient name, national identity card numbers called as direct identifiers. The attributes that are the individual's demographics and diagnosis codes are called as *Quasi-identifiers*. The attributes which are very sensitive to patients to hide details from others are called as sensitive identifiers.
In figure 4, the darkest red means there is a positive correlation in the datasets, while the darkest blue means there is a negative correlation. There is no correlation between 2 variables the color is gray. Figure 4 shows comparisons of attributes to find the sensitive attribute. Using the colors we can identify the sensitive attribute for the ranking method.

Figure 5 shows the feature selection method, which is by classifying datasets by Pearson correlation method to identify the sensitive attribute for the creation of hierarchy in data. The correlation values give data reliability.

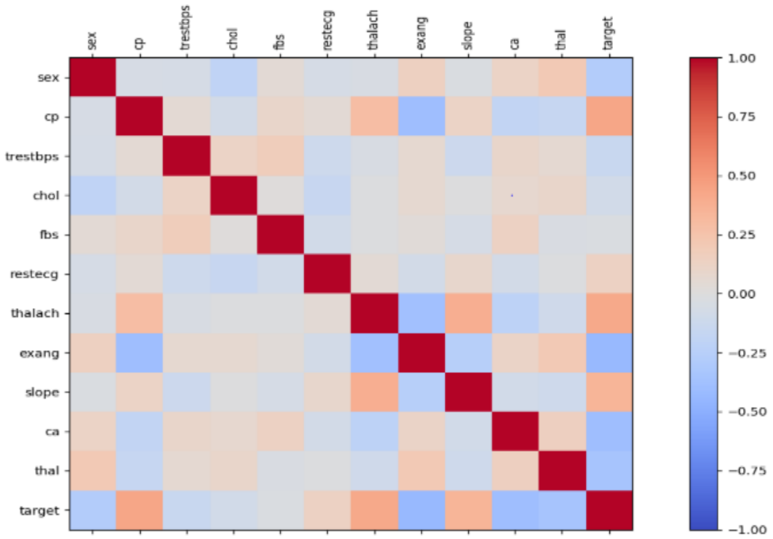


Figure 4: Correlation between the variables of the dataset

In Figure 5 the attribute-based classification is done by Pearson correlation method. The values with the highest priority are identified through the chart. Every attribute and its range of value are as shown in the above Pearson chart in figure 5.

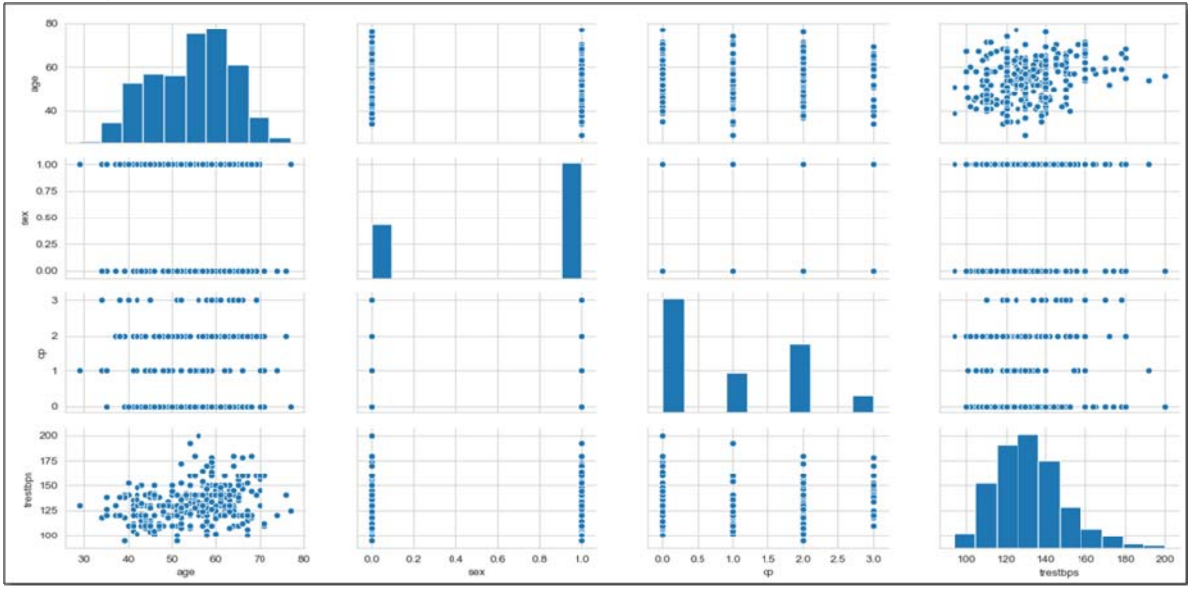


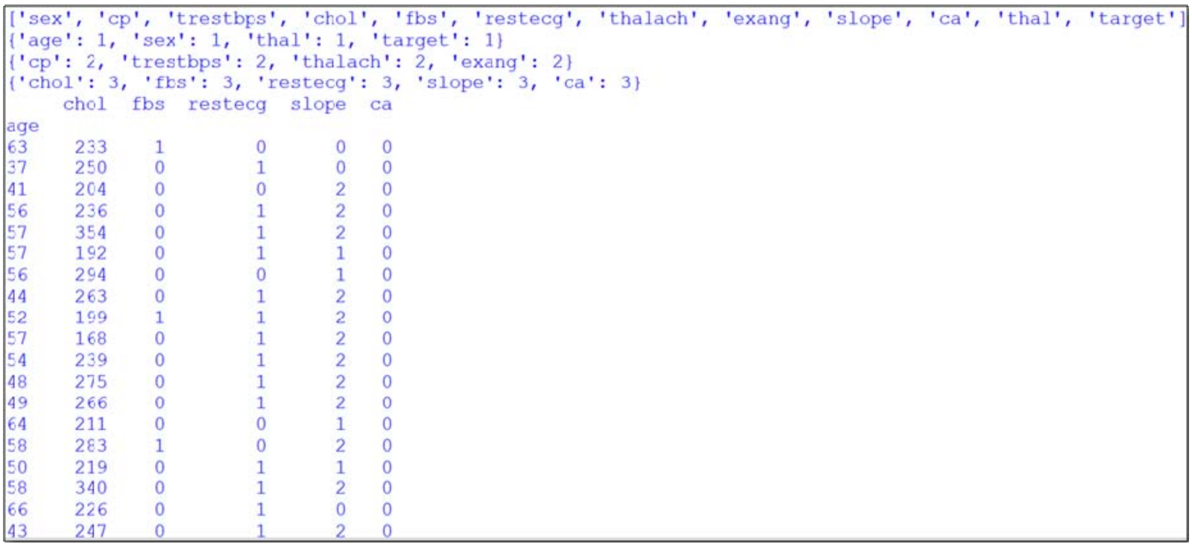
Figure 5: Feature selection (Pearson correlation)dataset comparison chart.

The Supervised algorithm is used for classification and regression problems. It analysis the datasets then create a conditional function for the training data by using supervised algorithms. There are two types of data, one is the training data and another one is the test data. Learning knowledge from the given input data is the training data, applying that knowledge from the training data to the test data for a newer output.

5. Decision Tree Algorithm for Creating Hierarchy

The decision tree's basic idea is to select the attribute by using Attribute Selection Measures to classify into different splits of the records. The best node is the decision node and splits the datasets into subsets below the decision node. Recursively doing the same to create subsets and to build the tree up to a condition satisfied. The conditions are; tuples all belong to the same attribute, there is no more balance in attributes and instances.

The Attribute Selection Method (ASM) delivers a rank to each attribute by explaining the given datasets. The source is the best attribute. The most popular attribute selection methods are the Information Gain method, Gini Index method and the Gain Ratio method.



```
['sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach', 'exang', 'slope', 'ca', 'thal', 'target']
{'age': 1, 'sex': 1, 'thal': 1, 'target': 1}
{'cp': 2, 'trestbps': 2, 'thalach': 2, 'exang': 2}
{'chol': 3, 'fbs': 3, 'restecg': 3, 'slope': 3, 'ca': 3}
     chol  fbs  restecg  slope  ca
age
63    233    1        0      0   0
37    250    0        1      0   0
41    204    0        0      2   0
56    236    0        1      2   0
57    354    0        1      2   0
57    192    0        1      1   0
56    294    0        0      1   0
44    263    0        1      2   0
52    199    1        1      2   0
57    168    0        1      2   0
54    239    0        1      2   0
48    275    0        1      2   0
49    266    0        1      2   0
64    211    0        0      1   0
58    283    1        0      2   0
50    219    0        1      1   0
58    340    0        1      2   0
66    226    0        1      0   0
43    247    0        1      2   0
```

Figure 6: Hierarchy creation of the dataset

6. Decision Tree Classifier Using Scikit Learn

Scikit-learn(SK-learn) is a package in python library, it holds modules like data splitting, preprocessing, feature selection, tuning, supervised and unsupervised learning algorithms. And other libraries used here are pandas, numpy, etc. Pandas for loading the dataset into a data frame, the data frame then passed as an input for the classifier, numpy numerical calculations in the analysis of data.Import required libraries to work with python. First load the datasets using pandas, then divide given file columns into two types of variables that are dependent and independent. For the understanding of the model, we must split the dataset into a training set and test set. To create a decision tree, create a tree classifier object using Scikit.

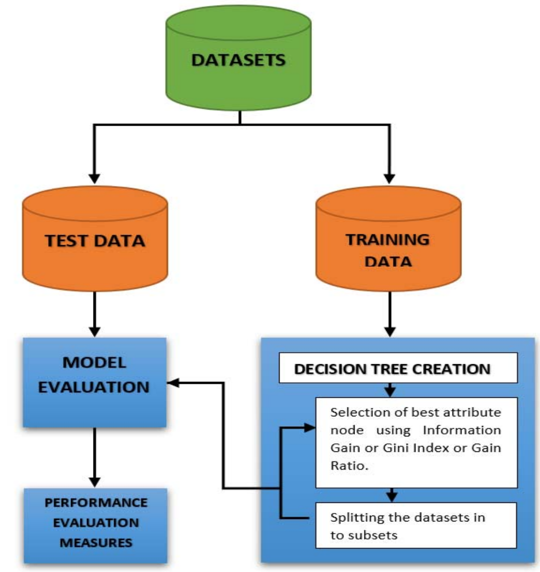

DATASETS
TEST DATA
TRAINING DATA
MODEL EVALUATION
PERFORMANCE EVALUATION MEASURES
DECISION TREE CREATION
Selection of best attribute node using Information Gain or Gini Index or Gain Ratio.
Splitting the datasets in to subsets

Figure 7: Dataset classification for the prediction

7. Conclusion

The sharing of electronic health records by paper works and by electronic access by different users affects patient's privacy and security. To overcome the individuals EHR sharing issues, the proposed system timing enabled designated distributed keys to different users in a hierarchical system. The datasets are classified into different subsets in a hierarchical pattern based on the sensitivity of the attribute. The higher priority attribute that is the attribute with sensitive information will act as a top node in the tree which is accessed by the higher grade users only. From the top-level node of the tree to the bottom node the sensitiveness of the attribute decreases. The user's hierarchy designed based on the designation, different grades having a different level of access. By hiding sensitive data to lower grade users with an access time limit in a hierarchy system will give high privacy and security.

References

[1] ArisGkoulalas, GrogoriousLoukides, Jimeng Sun, (2014). Publishing data from electronic health records while Preserving Privacy a Survey of Algorithms.*Journal of Biomedical informatics, Elsevier,* 50. 4-19

[2] C. Kalpana, S. Revathy (2018). A survey on policy combined and time-bound hierarchical delegation model in electronic health records.*Journal of Advanced Research in Dynamical and Control Systems*, 10(03-Special Issue). 853-855.

[3] C. Kalpana, S. Revathy (2017), Analysis of time-bound collaborative access control delegation model in electronic health records, *IEEEXplore Digital library*, DOI: 10.1109/ICCMC.2017.8282724

[4] Ivy Joy G. Mallare and Susan Pancho-Festin (2013). Combining Task- and Role-based Access Control with Multi-Constraints for a Medical Workflow System", *IEEE Conference Publications*, DOI: 10.1109/ICITCS.2013.6717814.

[5] Kwangsoo Seo, Young-gab Kim, Euijong Lee, Young-Duk Seo, and Doo-Kwon Baik (2018).Privacy-Preserving Attribute-Based Access Control Model for XML-Based Electronic Health Record System. *IEEE ACCESS,* 3.9114 - 9128

[6] LifengGuo and Wei-Chuen. Yau(2015), Efficient secure-channel free public key encryption with a keyword search for EMRs in cloud storage,*Journal of Medical Systems. Springer.* 39(2), 1–11.

[7] Mario Sicuranza, Angelo (2013). An Access Control Model for easy management of patient privacy in ERR systems". IEEE International Conference for Internet Technology and Secured Transactions, 2013.

[8] Maozhen Ding, FeiGao ; Zhengping Jin, Hua Zhang (2012). An efficient public key encryption with conjunctive keyword search scheme based on pairings,DOI**:** 10.1109/ICNIDC.2012.6418809

[9] Md. Zahangir Alam, M. Saifur Rahman, M. Sohel Rahman (2019). A Random Forest-based predictor for medical data classification using feature ranking. *Elsevier*, 15. 100-180

[10] M.-S. Hwang, S.-T. Hsu and C.-C. Lee (2014). A new public-key encryption with conjunctive field keyword search scheme,*Information Technology and Control*, 43(3). 277–288.

[11] Nafiseh Kahani, Canada Khalid Elgazzar, James R. Cordy, (2016) Authentication and Access Control in e-Health Systems in the Cloud. IEEE Published in Big Data Security on Cloud.DOI: 10.1109/BigDataSecurity-HPSC-IDS.2016.43.

[12] Po-Yen Wu, Chih-Wen Cheng, Chanchala D. Kaddi, JananiVenugopalan, Ryan Hoffman, and May D. Wang (2017). -Omic and Electronic Health Records Big Data Analytics for Precision Medicine, *IEEE Transactions on Biomedical Engineering*, 64(2). 263-273

[13] RandikeGajanayake, Renato Iannella NEHTA Brisbane, Tony Sahama(2012). Privacy Oriented Access Control for Electronic Health Records. ACM.

[14] Rui Zhang, Ling Liu and Rui Xue (2014), Role-Based and Time-Bound Access and Management of EHR Data, ACM journal Security and Communications Networks, 7(6).994-1015.

[15] Sebastian Haas, Sven Wohlgemuth, Isao Echizen, Noboru Sonehara, Günter Müller (2010). Aspects of privacy for electronic health records. *International Journal of Medical Informatics, Elsevier Ireland Ltd.* e26-e31

[16] Shalini Bhartiya, Deepti Mehrotra, Anup Girdhar (2017), Proposing hierarchy-similarity based access control framework: A multilevel Electronic Health Record data-sharing approach for the interoperable environment, Elsevier Journal of Computer and Information Sciences, 29(4). 505-519

[17] Sujatha .V, Prasanna Devi S, Vinu Kiran S, Manivannan S(2016), Bigdata analytics on Diabetic Retinopathy Study (DRS) on real-time data set identifying survival time and length of stay. *ELSEVIER*, 87. 227-232

[18] UthpalaPremarathne, AlsharifAbuadbba, AbdulatifAlabdulatif, Ibrahim Khalil, ZahirTari, Albert Zomaya, RajkumarBuyya(2016). Hybrid Cryptographic Access Control for Cloud-Based EHR Systems", IEEE cloud computing published by the IEEE Computer Society, 3(4).58-64

[19] Wei Li, Bonnie M. Liu, Dongxi Liu, Ren Ping Liu, Peishun Wang, Shoushan Luo and Wei Ni, Wei Li, Bonnie M. Liu, Dongxi Liu, Ren Ping Liu, Peishun Wang, Shoushan Luo, and Wei Ni,(2018). Unified Fine-Grained Access Control for Personal Health Records in Cloud Computing, *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS,* 5(3).1278 – 1289.

[20] XING GUANG ZHOU, JIAN WEI LIU, QIANHONG WU, AND ZONGYANG ZHANG, (2018). Privacy Preservation for Outsourced Medical Data with Flexible Access Control.*IEEE ACCESS*, 6.

[21] Xingguang Zhou, Jianwei Liu (2016), Anonymous role-based access control on e-health records. Proceedings of the ACM on Asia conference on computer and communications security, 03. 559-570.

[22] Yang Yang and Maode Ma, (2015). Conjunctive Keyword Search with Designated Tester and Timing Enabled Proxy Re-Encryption Function for E-Health Clouds.*IEEE Transactions on Information Forensics and Security*, 11(4).746-759.

[23] Yi Mu University of Wollongong (2016). Multi-Authority Security Framework for Scalable EHR Systems, *International Journal of Medical Engineering and Informatics,*8(4).390-408.

Authors Profile

C.Kalpana is an Research Scholar from the Department of Computer Science and Engineering in Sathyabama Institute of Science and Technology, Chennai India. Her research interest includes Artificial Intelligence, Machine Learning, Data Analytics and Big Data. She has published many publications in refereed journals.

Dr.S Revathy is an Associate Professor from the Department of Computer Science and Engineering working in Sathyabama Institute of Science and Technology, Chennai India. Her research interest includes Machine Learning, Data Analytics and Big Data. She has published more than thirty publications in refereed journals. She is the reviewer of many refereed journals.