PROVIDING PRIVACY AND SECURITY FOR CLOUD DATA USING CANFIS AND SCEA TECHNIQUE

R.Sandhya[1], M.Rekha[2]
[1,2]Computer Science and Engineering
Indira Institute of Engineering and Technology, Chennai, India

ABSTRACT

Cloud Computing is an environment for providing information and resources that are delivered as a service to end-users over the Internet on demand. Cloud storage service avoids the cost expensive on software, personnel maintenance and provides better performance, less storage cost and scalability. But the maintenance of stored data in a secure manner is not an easy task in cloud environment and especially that stored data may not be completely trustworthy. Cloud delivers services through internet which increases their exposure to storage security vulnerabilities. However security is one of the major drawbacks that prevent several large organizations to enter into cloud computing environment. This paper proposes two algorithms they are Secure Crypto Encryption and Co-Active Neural Fuzzy Interference System. SCEA propose for protect data access using efficient key management process. CANFIS propose to Data Mining for access upload file easy to make perfectly. Other words the usage of CANFIS is to process the file in a shortest path to access particular or needed files. In this paper we implement three important processes which are 1.Uploading bulk data securely, 2.Access or Modify huge file, 3.Transfer of large file with high security. A huge file is uploaded in the cloud storage and is encrypted by SCEA algorithm for security purpose by the file owner. The encryption key is sent to the owner mail id. If the user gives any input query, the related files are displayed based on the CANFIS mining algorithm. When the file is accessed by the user that time password sent to the registered mail id of the user by the cloud. Then the user details are sent to owner and the owner verify the user details and send the encryption key to decrypt the file. Finally the users access the file with a huge security over the cloud.

INTRODUCTION

DOMAIN INTRODUCTION

The project belongs to **"KNOWLEDGE & DATA MINING"** Approach

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), a relatively young and interdisciplinary field of computer science, is the process that results in the discovery of new patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract knowledge from an existing data set and transform it into a human-understandable structure for further use. Besides the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of found structures, visualization, and online updating.

The term is a buzzword, and is frequently misused to mean any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) but is also generalized to any kind of computer decision support system, including artificial intelligence, machine learning, and business intelligence. In the proper use of the word, the key term is discovery, commonly defined as "detecting something new". Even the popular book "Data mining: Practical machine learning tools and techniques with Java" (which covers mostly machine learning material) was originally to be

named just "Practical machine learning", and the term "data mining" was only added for marketing reasons. Often the more general terms "(large scale) data analysis", or "analytics" - or when referring to actual methods, artificial intelligence and machine learning - are more appropriate.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indexes. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting are part of the data mining step, but do belong to the overall KDD process as additional steps.

The related terms data dredging, data fishing, and data snooping refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations.

Overview

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

This white paper provides an introduction to the basic technologies of data mining. Examples of profitable applications illustrate its relevance to today's business environment as well as a basic description of how data warehouse architectures can evolve to deliver the value of data mining to end users.

The Foundations of Data Mining

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996.1 In some industries, such as retail, these numbers can be much larger. The accompanying need for improved computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Data mining algorithms embody techniques that have existed for at least

10 years, but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods.

In the evolution from business data to business information, each new step has built upon the previous one. For example, dynamic data access is critical for drill-through in data navigation applications, and the ability to store large databases is critical to data mining. From the user's point of view, the four steps listed in Table 1 were revolutionary because they allowed new business questions to be answered accurately and quickly.

The Scope of Data Mining

Data mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

• Automated prediction of trends and behaviors. Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

• Automated discovery of previously unknown patterns. Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.

1.2.1 PROJECT INTRODUCTION

In this section the clustering algorithms are used extensively only to organize and categorize file. In this paper proposed new clustering algorithms (CANFIS) are introduced. The algorithm introduces the most representative of line clustering techniques frequently used in conjunction with radial basis function networks and fuzzy modeling: subtractive clustering and introduces the most representative on line clustering techniques frequently called evolving clustering methods (ECM). The advantages of this method are 2-New fuzzy rules are created and updated during the operation of the system (through learning). Clustering partitions a data set into several groups such that the similarity within a group is larger than the among groups. Clustering techniques are used in conjunction with radial basis function networks or fuzzy modeling primarily to determine initial locations for radial basis functions or fuzzy if-then rules. For this purpose, clustering techniques are validated on the basis of the following assumptions. 1. States that the target system to be modeled is a smooth input outputs; mapping this is generally true for real world systems. 2. Requires the data set to conform to some specific type of distribution; however, this is not always true. Therefore, clustering techniques used for structure identification in neural or fuzzy modeling are highly heuristic, and finding a data set to which clustering techniques cannot be applied is not uncommon. If the user wants to upload his file to the cloud he first needs to get registered to his corresponding Cloud storage or server. The output of this registration process is the generation of a unique user identifier for that user by the KDC. This user ID will be further used for all operations being performed by the user in the cloud. This trustee system will generate a token for the user once he produces his unique Id to the trustee system. The generated token is further passed on to the corresponding KDC for generating the keys for encryption and decryption of the file that needs to be uploaded and/or downloaded. For secure file storage, a Homomorphic encryption technique is adopted which implements an asymmetric key cryptography called Secure Crypto encryption Algorithm, This Cryptosystem is

computationally strong and highly resistive to key-based attacks. It uses a series of complex mathematical functions for producing a single parameter. Now the files are encrypted using keys that are uniquely generated for this file according the access policy that has been defined by the owner of the file. The file is encrypted with keys that are generated by KDCs and also based on the access policy that is defined for that user by the owner of the file. Based on user authentication and claim policy, the files are encrypted and stored in the cloud. When some other user in cloud is interested in reading or writing the files, the access will be permitted based on the access policy defined for that particular user. Similar to the upload operation, the user sends a request for downloading the file from the cloud. He is authenticated and keys for decrypting the files are obtained using which the files are retrieved back.

SCOPE

Data mining is a powerful new technology, which aims at the extraction of hidden predictive information from large databases. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The process of knowledge discovery from databases necessitates fast and automatic clustering of very large datasets with several attributes of different types. This paper addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine grained data access control to un trusted cloud servers without disclosing the underlying data contents.

OBJECTIVE

- Apply SCEA for storage file in cloud based on generate encrypt and decrypt key and access securely by authentication user.
- Apply CANFIS for clustering process to query related data and grouped with display output query results.

SYSTEM ANALYSIS

PROBLEM DEFINITION

We live in a period of enormous information; its increasing amount has inserted a colossal potential and expanded complexity and risks such as data security as well as irrelevance and information overload. This information is stored in huge warehouses. Data or information mining is an alternate enhancement to help ventures to keep tabs on data in their warehouses. Cloud computing is an Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand. It is an upcoming paradigm that offers tremendous advantages in economical aspects, such as reduced time to market, flexible computing capabilities, and limitless computing power. Popularity of cloud computing is increasing day by day in distributed computing environment. There is a growing trend of using cloud environments for storage and data processing needs. To use the full potential of cloud computing, data is transferred, processed, retrieved and stored by external cloud providers. However, data owners are very sceptical to place their data outside their own control sphere. Their main concerns are the confidentiality, integrity, security and methods of mining the data from the cloud. Security becomes a major concern when using public cloud services.

EXISTING SYSTEM

Cloud computing provides the facility to access shared resources and common infrastructure, offering services on demand over the network to perform operations that meet changing business needs. With the help of ubiquitous internet technology, cloud providers offer numerous services to many sectors such as industry, academia, medical and the government. With the advent of cloud computing usage, adoption of cloud storage has become very simple. At the same time, providing data security becomes very momentous. The biggest impediment to the cloud environment is the deficiency of trust, which has stopped the shift of the whole of IT systems onto the cloud. Encryption is the genuine solution to mitigate data security risks. But cloud service providers do not support encrypted data. Privacy preserving data mining ensuring the security of individual data or precise knowledge without relinquishing the utility of the data. Individuals have ended up very much aware of the privacy interruptions on their own data and are extremely

hesitate to import their sensitive data. This may prompt the unintentional results of data mining.

ISSUES IN EXISTING SYSTEM

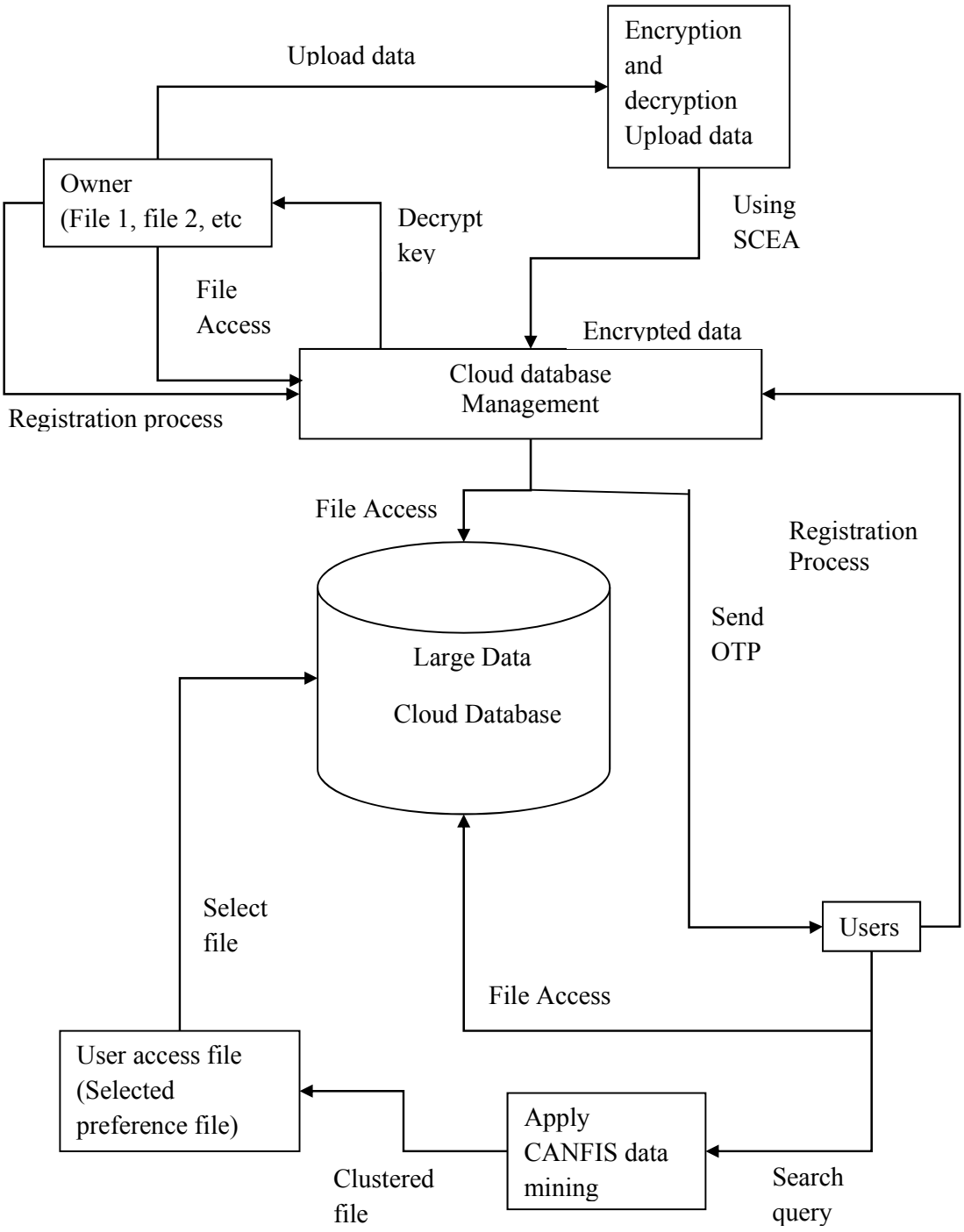
- Result is not good
- High processing time
- Need Security & Reliability result

PROPOSED SYSTEM

In this paper propose two algorithm techniques are SCEA and CANFIS, SCEA implement for security purpose and CANFIS implement for file accessing process i.e quickly get user preference files in hybrid cloud. In this paper important another processing is registration between owner and user. This paper main object is registration details, file access and upload based working on registration object values. SCEA Implement for encryption process for upload files in hybrid. Upload files before encrypted using crypto technique process, SCEA using three important processes are key generation, encryption and decryption. CANFIS using only clustering process. First registration process completes after upload a file using crypto technique process. If any user access that file using registration key for security purpose allow access hybrid database. Users search a file because user preference files to given text type. In this time using CANFIS algorithm for clustered all files and display or given output of user's preference file with show all related files. Users select that file at a time generate OTP by cloud and send to mail id for verify authenticate or file selected users. Finally, users receive OTP to open selected or preference file securely.

ADVANTAGES

- Processing time is low
- Improve the quality result
- High security
- Accuracy & scalability results
- Given output is reliability

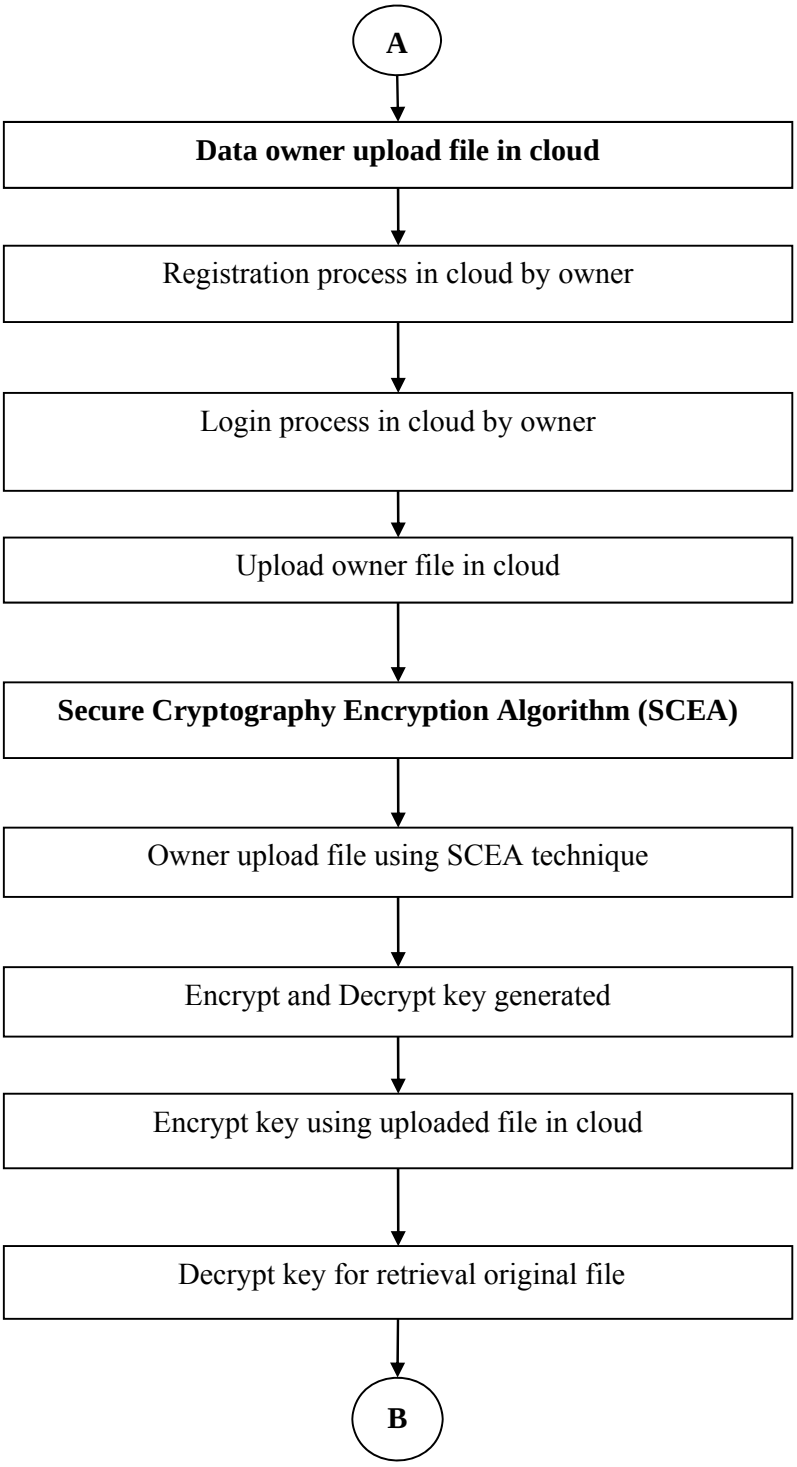


SYSTEM DESIGN & IMPLEMENTATION

SYSTEM ARCHITECTURE

Fig: overall proposed architecture

SYSTEMS DESIGN



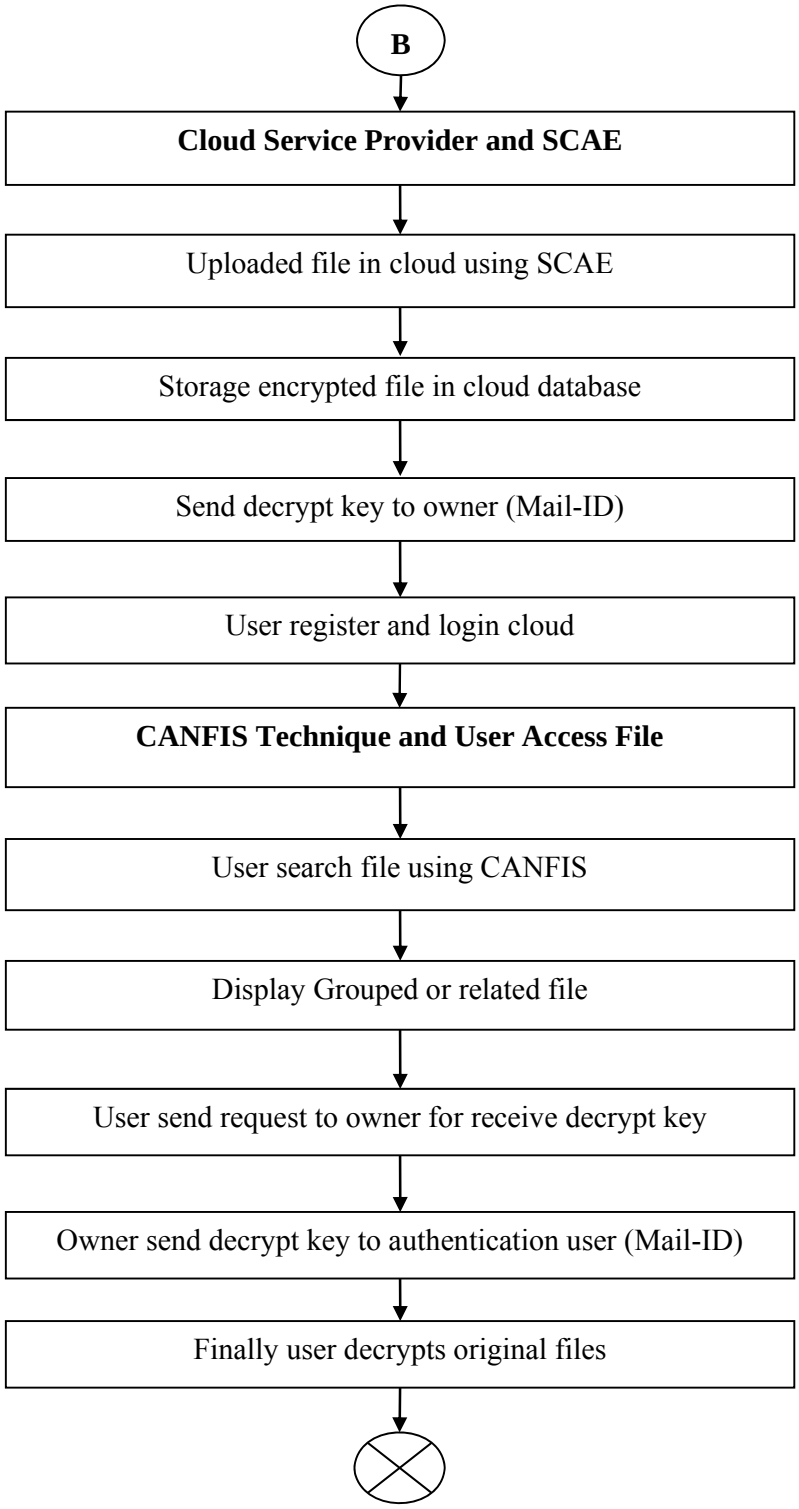


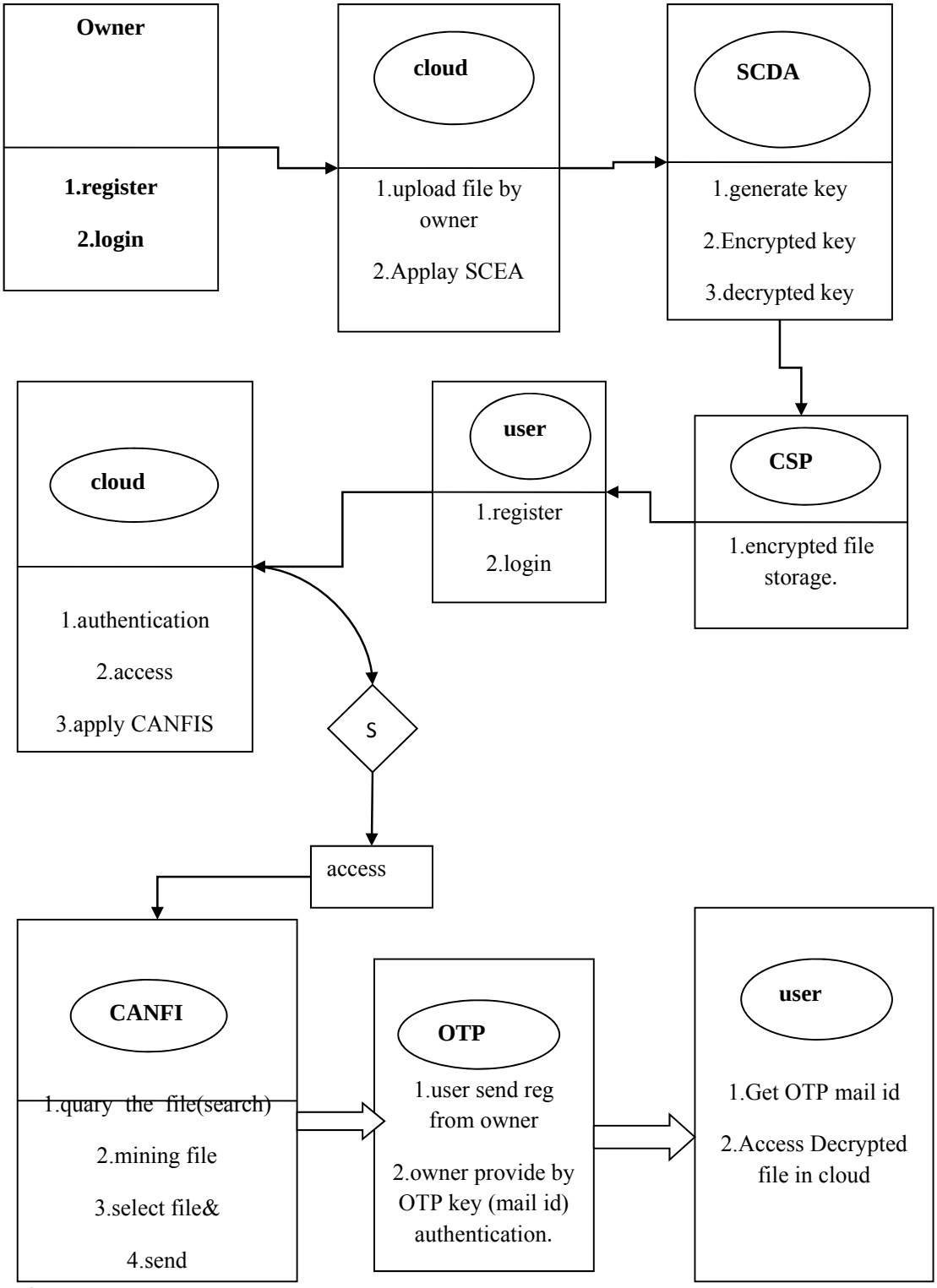
Fig: Implementation deign

DATAFLOW DIAGRAM

The Data Flow diagram is a graphic tool used for expressing system requirements in a graphical form. The DFD also known as the "bubble chart" has the purpose of clarifying system requirements and identifying major transformations that to become program in system design. Thus DFD can be stated as the starting point of the design phase that functionally decomposes the requirements specifications down to the lowest level of detail. The DFD consist of series of bubbles joined by lines. The bubbles represent data transformations and the lines represent data flows in the system. A DFD describes what that data flow in rather than how they are processed. So it does not

depend on hardware, software, data structure or file organization.

Data flow diagrams are used to describe how the system transforms information. They define how information is processed and stored and identify how the information flows through the processes. A DFD provides no information about the timing of processes, or about whether processes will operate in sequence or in parallel. It is therefore quite different from a flowchart.



Dataflow diagram

UML DIAGRAM

ACTIVITY DIAGRAM

Activity Diagrams describes about the sequences of the activities in a system with the help of Activities. An Activity is a single step in a process. One Activity is one state in the system with internal activity and atleast one outgoing transition. Activities can also have more than one outgoing transition if they have different conditions.

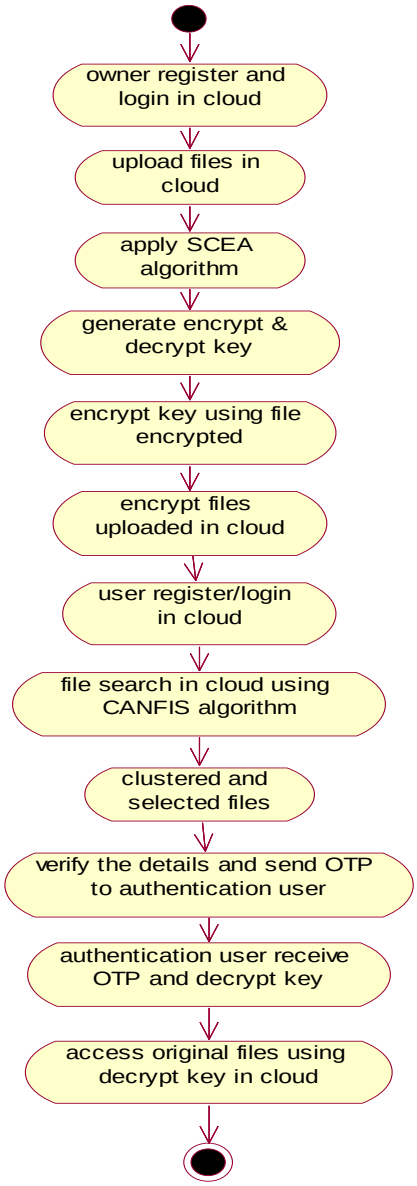


Fig: Activity diagram

USE CASE DIAGRAM

Use Case Diagrams describes the relationships and the dependencies between a group of Use Cases and the Actors participating in the process. Use Case Diagrams are meant to facilitate the communication with the future users of the system and with the customer and are especially helpful to determine the required features of the system is to have. Use Case Diagrams describes what the system should do but do not and cannot specify how this is to be achieved.

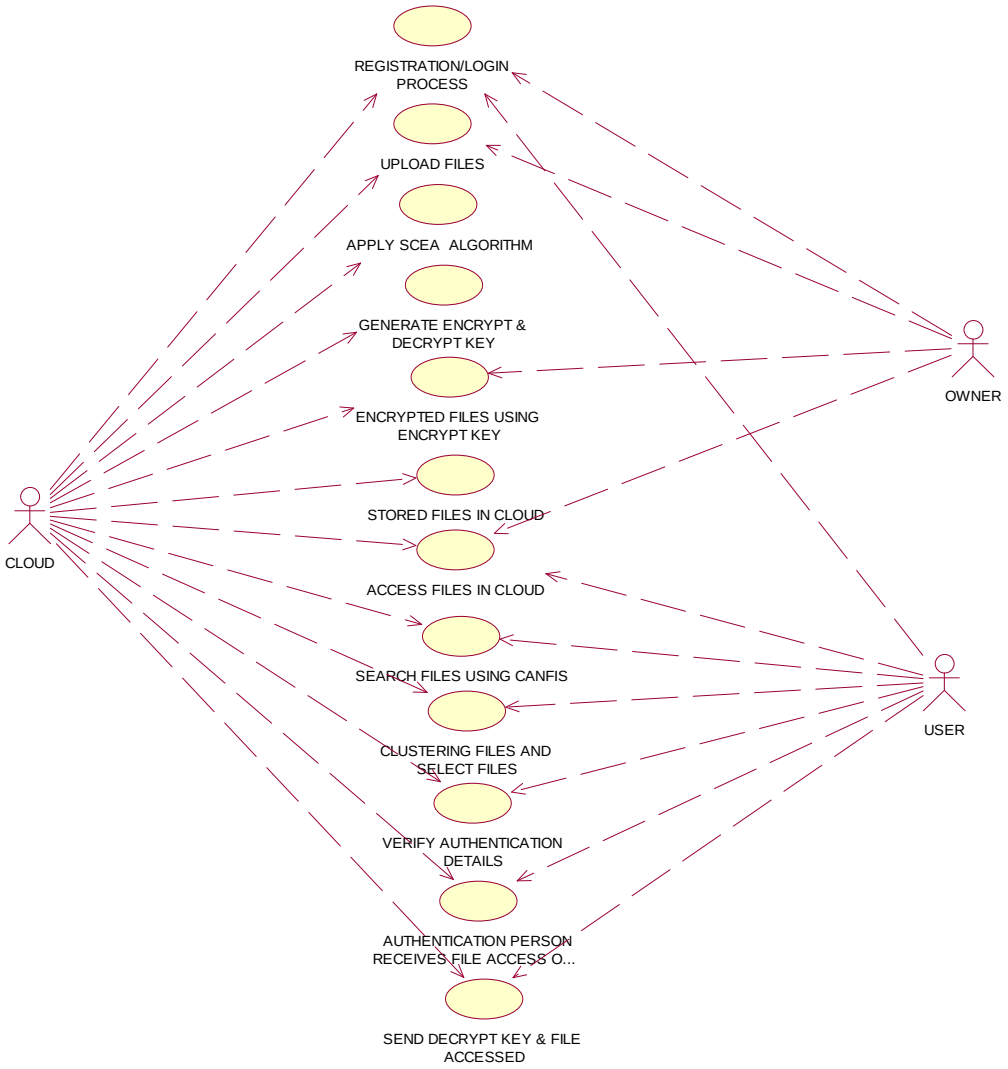


Fig: Use case diagram

COLLABORATION DIAGRAM

Collaboration Diagram depicts how collaborations are wired together to form larger collaborations between the entities. They are used to illustrate the structure of arbitrarily complex systems. Collaborations are wired together by using an assembly connector with number between the connections.

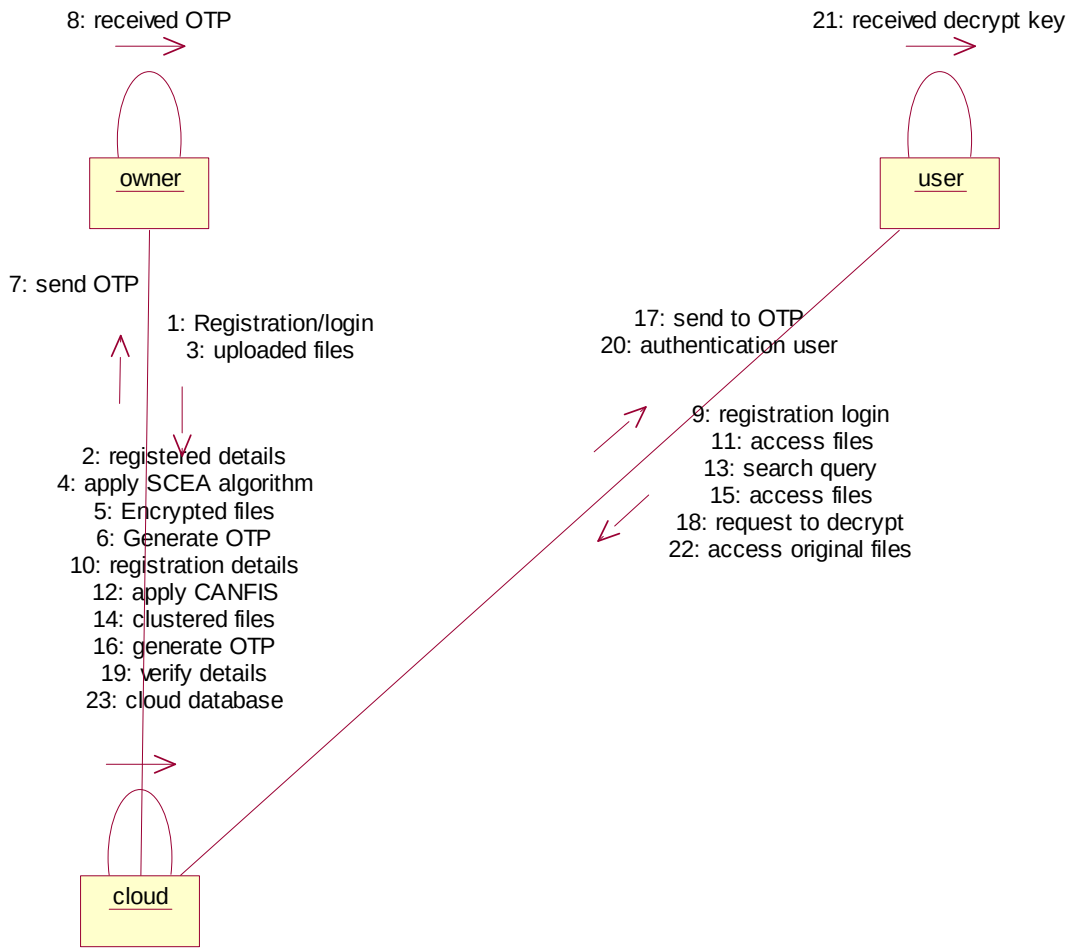


Fig: collaboration diagram

SEQUENTIAL DIAGRAM

Sequence Diagrams shows the message exchange (i.e. method call) between the several objects in a specific time-delimited situation. Sequence Diagrams put special emphasis in the order and the times in which the messages to the objects are sent. The following Diagrams explain about sequence process.

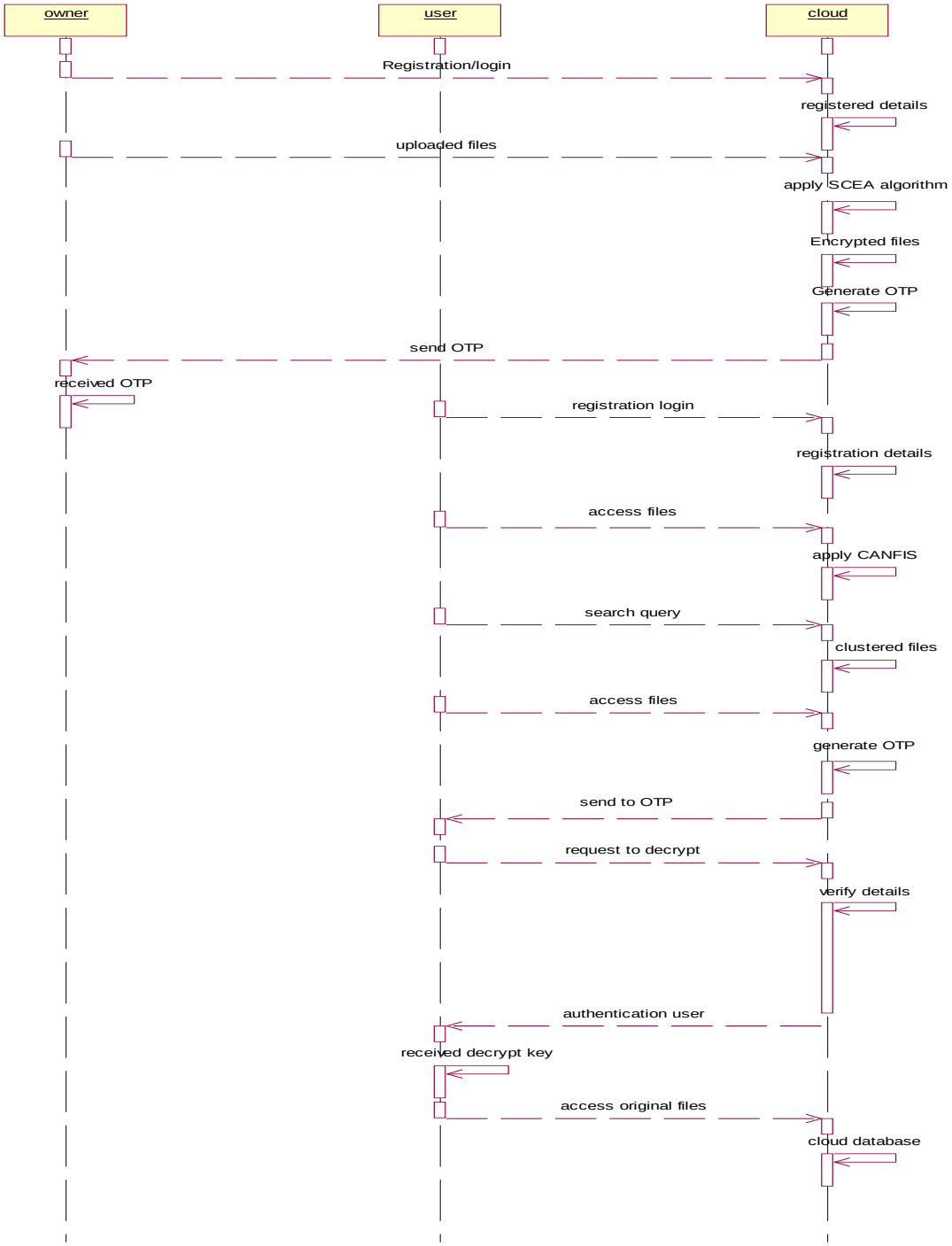


Fig Sequential Diagram

CLASS DIAGRAM

The Class Diagram is the main building bock of object oriented modeling. It is used for both general conceptual modeling of the systematic of the application and also for the detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagrams these classes are represented with the boxes which contain three parts.

- The upper part holds the name of the class
- The middle part contains the attributes of the class

The bottom part gives the methods or operations of the class which can take or undertake.

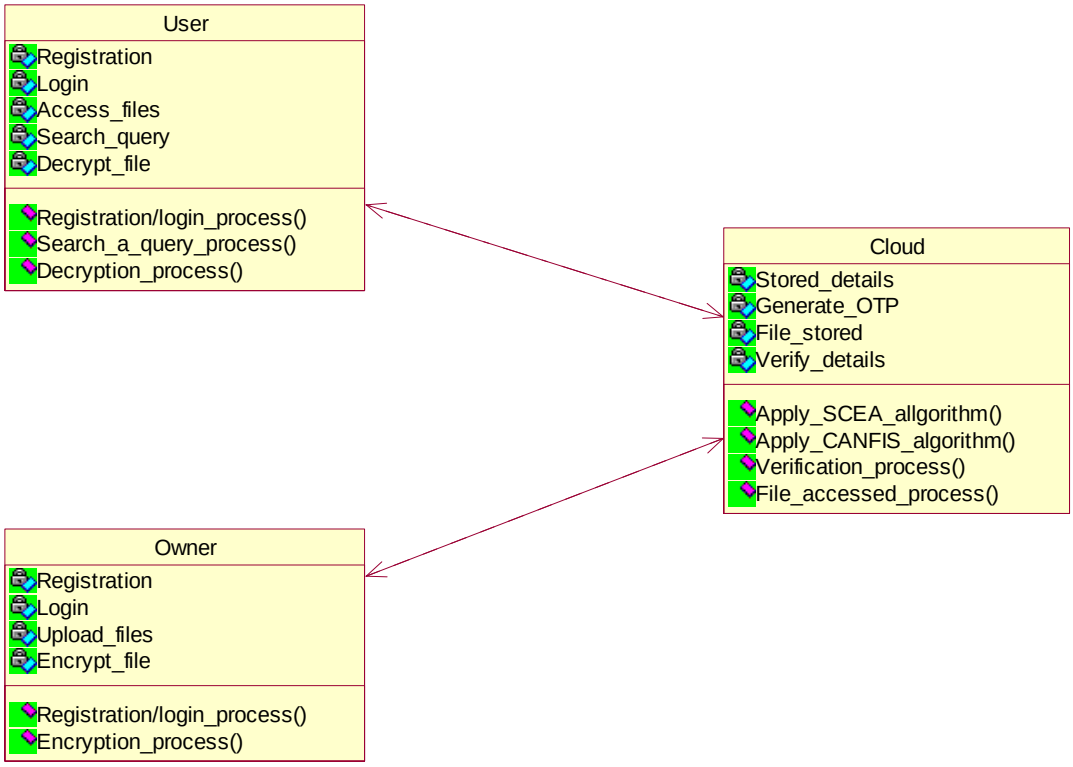


Fig: Class diagram

SYSTEM IMPLEMENTATION

FUNCTIONAL MODULES

- File Encryption and Upload:
- Query process
- Fuzzy clustering process
- User Access files by OTP

MODULE EXPLANATION

File Encryption and Upload:

The secure crypto encryption system is implemented here for generating keys. The users now encrypt their files with the keys received keys. They also set their own access policies, i.e. privileges to the file. The access policies set by individual users for their files are hidden from other users by implementing a query driven approach. The query-driven approach is an SQL coding written for the cloud database. Here in, the attributes and privileges of users are hidden from the cloud as their details are stored in encrypted format.

Query process

Owner and User both access uploaded a file in hybrid cloud storage space. This process basically called recommendation method type of fuzzy logic system. User give query of file name in cloud, CANFIS system to be clustered (partition or classification) and got user given query input files. That files mostly users liked or recently update file shown in database system.

Fuzzy clustering process

In our proposed system we use fuzzy logic system recommendation technique. It is a problem solving control of system methodology. It lends itself for implementation systems ranging from small, simple, embedded of micro-controllers to large, multi channel PC or workstation based data acquisition, networked and control systems. FL can be implementing in software, hardware or both platforms. Fuzzy logic provides easiest way to come at a certain conclusion based on vague, imprecise, noisy, ambiguous, or missing input of information. Fuzzy logic approach is used to control a

problem and helps how a person would make decision.

User Access files by OTP

Some process policy following implemented in user side for security level increased and identify who known access that file as user details. After search a file name (give query) in cloud engine. In this time using CANFIS algorithm for clustered large data files and give user preference file quickly. User select at a time cloud generate OTP for verifies the user detail after send OTP to user mail id. So finally, user mail id open and access upload file in cloud database by user.

ALGORITHM USED

SECURE CRYPTO ENCRYPTION ALGORITHM:

A) Key Generation

1) Choose two large prime numbers p and q, such that gcd (pq, (p-1) (q-1)) = 1.
2) Compute n = pq, λ = lcm (p-1, q-1).
3) Select random integer g such that $g \in Z * n^2$
4) Calculate the following modular multiplicative inverse
5) $\mu = (L (g^{\lambda} \bmod n^2))^{-1} \bmod n$, where the function L is defined as L(u) = u – 1/n.
6) The public (encryption) key is (n, g).
7) The private (decryption) key is (λ,).

B) File Encryption

1) Let m be a message to be encrypted where m $\in Z_n$.
2) Select random r where r ∈ Z*.
3) Compute cipher text as, $c = g^{m} \cdot r^{m} \bmod n^2$

C) File Decryption

1) Cipher text $c \in Z * n^2$
2) Compute message, $m = L (c^{\lambda} \bmod n^2) \mu \bmod n$

CO-ACTIVE NEURO FUZZY INFERENCE SYSTEM ALGORITHM

Layer 1: Each node in this layer performs fuzzification and generates membership grade of linguistic label of an input variable.

Layer 2: Each node in this layer is denoted by determining the MF of the whole input vector by aggregating the fuzzified results of the individual scalar functions of the every input variable. The output of each node in this layer is obtained by multiplying the incoming signals and represents the firing strength of a rule.

Layer 3: Each node in this layer is labeled as N and computes the normalized firing strength.

Layer 4: The output of each node in the fourth layer is calculated by the sum of the signals of the third and second layer of the network.

Layer 5: There is only single node in this layer labeled as Σ that calculates the overall output of the ANFIS or CANFIS as the summation of all incoming signals.

CONCLUSIONS

This paper proposes to secure data upload and access securely with retrieval file make access easily to handle in hybrid cloud database. it using two algorithm for secure with early access file, in this paper propose to Secure Crypto Encryption Algorithm and Co-Active Neuro Fuzzy Interference System. CANFIS modeling was proposed as a dependable and robust method developed to identify a nonlinear relationship and mapping between the different attributes. It has been a very useful technique for auto-tuning of the CANFIS parameters and selection of optimal feature set. The efficiency of the system can be analyzed in terms of encryption and decryption time of the algorithm. We compare the performance of the system with a symmetric key encryption (3DES) system. We can see that the proposed system SCEA is at par with performance when compared to symmetric key based system when files of different size were given as input and encrypted. The results show the encryption and decryption time is fast when compared with the symmetric algorithm based system. Thus the proposed system is fast and secure in terms of file recovery and file encryption. The time required to find the files been corrupted is also fast and the recovery of corrupted files back to original takes reduced amount of time.

REFERENCES

[1] B. Sevak, "Security against side channel attack in cloud computing," International Journal of Engineering and Advanced Technology (IJEAT), vol. 2, no. 2, p. 183, 2013.

[2] Study of Data Mining algorithm in cloud computing using MapReduce Framework Viki Patil, Prof. V. B. Nikam Journal of Engineering, Computers &

Applied Sciences (JEC&AS) ISSN No: 2319-5606 Volume 2, No.7, July 2013

[3] S. Ruj, M. Stojmenovic and A. Nayak, "Decentralized Access Control with Anonymous Authentication of Data Stored in Clouds", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 2, pp. 556-563, 2013.

[4] A New Algorithm to Model Highly Nonlinear System based Coactive Neuro Fuzzy Inference System Tharwat O. S. Hanafy International Journal of Computer Applications (0975 – 8887) Volume 94 – No.17, May 2014

[5] Information Systems in Management (2015) Vol. 4 (4) 264-275 Systems And Models Of Artificial Intelligence In The Management Of Modern Organisations Mariusz Maleszak, Piotr Zaskórski.

[6] Wang, Z., K. Sha, and W. Lv, Slight Homomorphic Signature for Access Controlling in Cloud Computing. Wireless Personal Communications, 2013. 73(1): pp. 51-61

[7] Wang, Z., G. Sun, and D. Chen, A new definition of homomorphic signature for identity management in mobile cloud computing. Journal of Computer and System Sciences, 2014. 80(3): pp. 546-553

[8] Li, H., et al., Identity-Based Authentication for Cloud Computing, in Proceedings of the 1st International Conference on Cloud Computing. 2009, Springer-Verlag: Beijing, China. pp. 157-166

[9] Qin, B., et al., Simultaneous authentication and secrecy in identity-based data upload to cloud. Cluster Computing, 2013. 16(4): pp. 845-859

[10] Mishra, D., V. Kumar, and S. Mukhopadhyay, A Pairing-Free Identity Based Authentication Framework for Cloud Computing, in Network and System Security, 2013, Springer Berlin Heidelberg. pp. 721-727

[11] Acar, T., M. Belenkiy, and A. Küpçü, Single password authentication. Computer Networks, 2013. 57(13): pp. 2597-2614

[12] Dinesha, H.A. and V.K. Agrawal. Multi-level authentication technique for accessing cloud services. International Conference on Computing, Communication and Applications (ICCCA), 2012

[13] Yassin, A., et al., Cloud Authentication Based on Anonymous One-Time Password, in Ubiquitous Information Technologies and Applications, 2013, Springer Netherlands. pp. 423-431

[14] Abdellaoui, A., Y.I. Khamlichi, and H. Chaoui, A Novel Strong Password Generator for Improving Cloud Authentication. Procedia Computer Science, 2016. 85: pp. 293-300

[15] Celesti, A., et al. Three-Phase Cross-Cloud Federation Model: The Cloud SSO Authentication. in Advances in Future Internet (AFIN), 2010 Second International Conference on. 2010.