IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Anonymous Authentication of Data Stored on Distributed Cloud

Prof. Sagar Rajebhosale[1], Mr. Pravin Dobe[2], Mr. Piyush Dani[3], Mr. Kaustubh Kulkarni[4], Mr. Sunil Jadhav[5]

Assistant Professor, Dept of Information Technology, PVG's College of Engineering, Nashik, India[1]

UG Student, Dept of Information Technology, PVG's College of Engineering, Nashik, India[2, 3, 4, 5]

Abstract: Authors propose an anonymous authentication for secure data storage in clouds for different departmental activities of an institute. In this scheme, the cloud verifies the authenticity of the user without knowing the users identity before storing data, also has the feature of access control in which only valid users are able to decrypt the stored information. Along with access control user revocation is also done that is existing user can be removed and cannot further access the data stored in cloud. The cloud storage is distributed. The cloud storage can distributed according to the requirement of the organization. System is very useful for sending messages anonymously. But only authorized user can do so i.e. anonymous but valid user only can send messages. This makes cloud computing a more secure approach. The scheme also gives the feature of text filtration which eliminates the meaningless words to optimize cloud storage. Moreover, our scheme gives the feedback policy needed for the performance improvement with different parameters anonymously. The main focus is on anonymous authentication so that one can give valid feedback, complains without revealing once identity. Authors aim to promote paperless work by means of e-notices. This approach can prove helpful for government and non-government organizations.

Keywords: Access Control, Authentication, Attribute-Based Signatures, Attribute Based Encryption, Cloud Storage, Text Filtering.

1. INTRODUCTION

In today's technological world, organizations are becoming more and more dependent on their information systems. The public is greatly concerned about the proper use of information, especially personal data. The threats to information systems from criminals and hackers are increasing. Many organizations will identify information as an area of their operation that needs protection from external and invalid users as part of their system of internal control.

More and more organizations are moving towards cloud storage rather than traditional schemes. Cloud has been boon to the organizations as the servers and its maintenance is looked after by cloud vendors. Cloud is fairly secure.

But cloud computing poses privacy concerns because the cloud service provider can access and manipulate the data that is on the cloud at any time. It could accidently or purposefully alter or even delete the data. Many cloud providers can share information with third parties either for purposes of law and order or for the purpose of advertisement even without a warrant or permission. These things are permitted in their privacy policies where every user have to accept it before actually start using the various cloud services. Solutions to privacy include lawmaking and policy as well as end users' choices for how data is stored. So, authors propose a decentralized and distributed scheme for cloud storage which can potentially cope up with the above issues. Decentralized scheme makes the system robust as single point of failure is bottleneck in centralized systems. Distributed cloud storage can reduce the overheads in significant amount. A separate cloud for every single department in an organization can be made. Our system anonymously authenticates user i.e. checks for valid user without revealing its identity. Then the user is able to store that data on cloud. For security purpose, the data is encrypted before storing it on cloud. The keys are distributed to those users whom the data owner has given the rights to view or modify the data. Access control restricts the unauthorized users from viewing or modifying the data.

In order to prevent the use of invalid language in the message text filtering mechanism is incorporated. This will disallow use of improper words and text in the message. Message integrity will be user related issue i.e. depending upon the user. The users will be able to upload, download and view files and messages on cloud if they have the proper access right and key to do so. Sharing of message can be done using this system. Complaints can be posted without revealing the identity. This all will promote paperless work.

2. LITERATURE SURVEY

In order to deal with vulnerabilities regarding data stored in cloud and its access policies, various schemes were proposed.

In 2006 A. Sahai and B. Waters, proposed Fuzzy Identity-Based Encryption. In Identity Based Encryption technique, users are given a set of attributes or properties along with its unique ID. A Fuzzy IBE scheme can be used for encryption. In Fuzzy scheme biometric inputs are used as identity. The advantage of that scheme was it was error tolerant. It was proved to be secured against collusion attacks. [1]

DOI 10.17148/IJARCCE.2016.53117

Another scheme named as Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data was introduced same year. In that scheme, the sender has an authorization to encrypt information. A revoked attributes and keys of users cannot write again to false or stale information. The attribute authority receives attributes and secret keys from the receiver and he/she is able to decode or decrypt data if it has matching attributes. The main advent of that scheme was distribution of audit log information was present. [2]

Improvements kept coming after that scheme. Cipher text-Policy Attribute-Based Encryption was one of them. By using this technique, the receiver has the access policy in the form of a tree. The tree contained attributes as leaves and monotonic access structure with the threshold gates like AND, OR and others. Advantages were encrypted information could be kept confidential even if the storage server was untrusted. Plus, it was secure against collusion attacks. [3]

In 2007 M. Chase proposed Multi-Authority Attribute Based Encryption. This scheme describes several Key Distribution Authorities (which are coordinated by a authority like trusted third party or trusted authority) which distribute attributes and secret keys to users. Multi authority Attribute Based Encryption protocol which needs no trusted authority which requires each user to have attributes from at all the KDCs. The fact that it allowed more number of attributes was the biggest advantage. [4] Decentralizing Attribute-Based Encryption scheme came to the picture. The scheme where users could have zero or more number of attributes from each authority and did not require a trusted server. Collusion resistant was there which proved as an advantage. [5]

In 2011 M. Green, S. Hohenberger, and B. Waters, proposed Outsourcing the Decryption of ABE Ciphertexts. In this paper the decryption task is done by a proxy Server, so that the user made computation on minimum resources like hand held devices. The user being able to significantly save bandwidth, without raising the number of transmission was biggest advantage. [6]

There also exists a decentralized access control scheme for secure data storage in cloud that supports anonymous authentication. In that scheme, the cloud verifies the authenticity of the user without knowing its identity before storing data. It has the feature of access control in which only valid users are able to decrypt the stored information. The scheme prevents replay attacks. Creation, modification, and reading data stored in the cloud are facilitated.

Moreover, authentication and access control technique is decentralized and robust, unlike other access control schemes designed for clouds which are centralized.

The scheme not only provides fine-grained access control but also authenticates users who store information in the cloud. The cloud however does not know the identity of the user who stores information, but only verifies the user's credentials. Key distribution is done in a decentralized way. [7]

3. PROPOSED SYSTEM

In the proposed system, the user will first login to the system. The login credentials will be verified by the trustee server. On successful login i.e. valid credentials, the user will get a token. Token is randomly generated value based on user credentials. Token is generated by the trustee server if and only if the credentials are valid and correct. For every user, the token generated will be different. Then the user will be able to upload and download the files and messages or complaints as shown in the figure 1. The file or message will be encrypted and decrypted using the cipher policy attribute based encryption algorithm.

The key generation will be handled by the KDC i.e. Key Distribution Center and the access rights will be stored on the same. KDC is a server for key generation and distribution. Multiple KDC's will be used in order to avoid bottle neck of failure of centralized KDC.

The users will only be able to encrypt and upload data if it is in vague free language i.e. does not contains any bad words. Text filtering using naïve bayes classifier will be done on the data. If and only if the message is proper without any slag words then it will be encrypted. The text filtering is limited only upto English language.

After the process of encryption and assigning access rights, the user will be able to upload the data on the cloud.

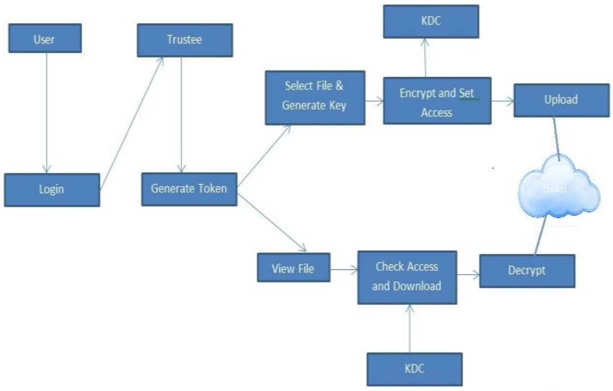


Figure 1:-Proposed System Architecture

While Decrypting a file, the access rights of the user will be checked if the rights are matching then the key will be given for decryption. After decryption, the user will be able to download and view the data.

Authors are designing the system for a typical college system. The hierarchy of the proposed system is shown in figure 2. In the system hierarchy, at the top level there is a administrator or super user i.e. principal. The admin will have access of all the departmental clouds. The admin will be able to add URLs of the KDCs, add HODs and will have control over trustee as well.

Head of Departments (HOD) will be the subordinate to the principal and will have full access for the respective departmental cloud. HOD will be able to add, revoke, update a user with the access rights of upload, download and view message. The user may be a student or staff.

Staff members will be subordinates of HOD. Staff will be

DOI 10.17148/IJARCCE.2016.53117

able to add, remove and update the user whose is a student. Staff will also be able to upload, download data and view the data which they have access for.

Student will be at the lowest level of hierarchy. Student will be able to upload, download and view data. Also, student can give feedback about college facilities and the staff members.

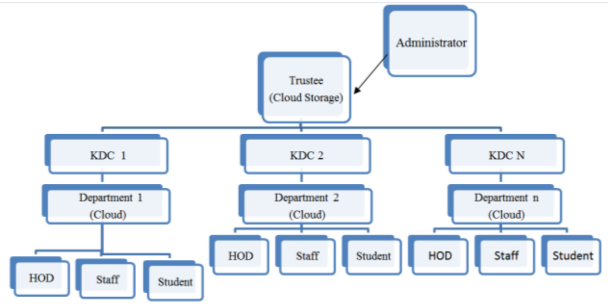


Figure 2:-Proposed System Hierarchy

3.1 Cipher Policy Based Attribute Based Encryption (CP-ABE)

In a CP-ABE scheme, every ciphertext is associated with an access policy on attributes, and every user's private key is associated with a set of attributes. A user is able to decrypt a ciphertext only if the set of attributes associated with the user's private key satisfies the access policy associated with the ciphertext. The access structure built in the encrypted data can let the encrypted data choose which key can recover the data; it means the user's key along with its attributes satisfies the access structure specified of the encrypted data. The data owner or encryptor specifies the access structure for his interested attributes while encrypting a message. Based on this access structure message is then encrypted such that only those whose attributes satisfy the access structure can decrypt it.

Access rights are assigned while encryption of the data. Access rights means that who can decrypt the data.

CP-ABE scheme consists of following four algorithms:

Setup: This algorithm takes as input a security parameter K and returns the public key PK as well as a system master secret key MK. PK is used by message senders for encryption. MK is used to generate user secret keys and is known only to the authority.

Encrypt: This algorithm takes as input the public parameter which contains user attributes and access structure, a token, filename and an access rights. It outputs the ciphertext CT.

Key-Gen: This algorithm takes as input a set of attributes associated with the user and access structure as parameters, a token, access rights and filename. It outputs a secret key SK that enables the user to decrypt a message encrypted under an access tree structure T if and only if matches T.

Decrypt: This algorithm takes as input the ciphertext CT and a secret key SK for an attributes set. It returns the message M if and only if satisfies the access structure associated with the ciphertext CT.

It can support the access control in the real environment. In addition, the user's private key in this scheme, is a combination of a set of attributes, so an user only use this set of attributes to satisfy the access structure in the encrypted data.[3]

3.2 Comparison of Various ABE Schemes

The comparison of various attribute based encryption schemes is presented in table 1 with respect to following parameters:

Fine Grained Access Control: Facilitates granting differential access rights to a set of users allowing flexibility specifying the access rights of individual users. In the same group, granting different access rights to individual users.

Efficiency: In terms of security.

Computational Overheads: Complexity of encryption algorithm.

Collusion Resistance: No two users can collude and access data or authenticate themselves, if they are individually not authorized. [9]

3.3 Naïve Bayes Classifier

Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes' Theorem with sturdy independence assumptions. It is an independent feature model. These independence assumptions of features make the features order is irrelevant and consequently that the present of one feature does not affect other features in classification tasks. These assumptions make the computation of Bayesian classification approach more efficient, but this assumption severely limits its applicability.

Contingent upon the exact way of the likelihood display, the Bayes classifiers can be prepared exceptionally well by requiring a generally little measure of preparing information to gauge the parameters essential for characterization. With very limited training data, the classifier can be trained well. Since independent variables are accepted, just the changes of the variables for every class should be resolved and not the whole covariance matrix. Because of its obviously over-rearranged suspicions, the Naïve Bayes classifiers regularly work vastly improved in numerous complex real world problems than one may anticipate. The Naïve Bayes classifier has been accounted for to perform well for some real world classification applications under some particular conditions. Favorable position of the Naïve Bayes classifier is that it requires a little measure of preparing information to measure the parameters important for order. Bayesian grouping approach touches base at the right order the length of the right class is more probable than the others. The classifier is robust as we can ignore serious deficiencies in its underlying naïve probability model. Class' probabilities don't need to be evaluated exceptionally well. Naïve Bayes functions admirably on numeric and textual data, simple to execute and calculation contrasting and different calculations, however works poorly when components are exceedingly related and does not consider recurrence of word events.[10][11]

DOI 10.17148/IJARCCE.2016.53117

3.4 Comparison of Text Classification Algorithms

Comparison of various text classification algorithms is given in the table 2. [11]

Table 1: Comparison of Various ABE Schemes

Technique/ Parameter	ABE	KP-ABE	CP-ABE	HABE
Fine Grained Access Control	Low	Low. High if there is re-encryption	Average Realization of complex Access Control	Good Access Control
Efficiency	Average	Average. High for broadcast system	Average	Flexible
Computational Overhead	High	Most of computation overheads	Average computation overhead	Some of overhead
Collusion Resistant	Average	Good	Good	Good

Table 2: Comparison of Text Classification Algorithms

Classification Algorithm	Advantages	Disadvantages
K-Nearest Neighbor	Simple and effective and easy to implement.	Hard to find out the value of K, time cost is more
Naïve Bayes	Easy for implementation and computation.	very poor when features are co related to each other
Support Vector Machine	Compact description of the learned model, more capable to solve multi-label classification	Training speed is slow
Neural Network	Provide better result in complex domain	Long training process
Decision Tree	Simple even non expert user can understand	Irrelevant attributes may affect badly the construction of a decision tree

4. PARTIAL RESULTS

Here are the snapshots of the system which is being implemented.

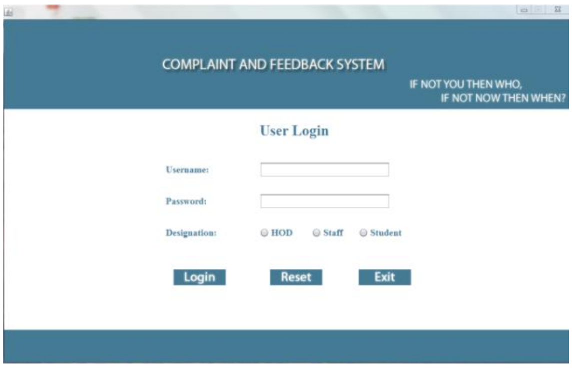


Figure 3:-Login Page

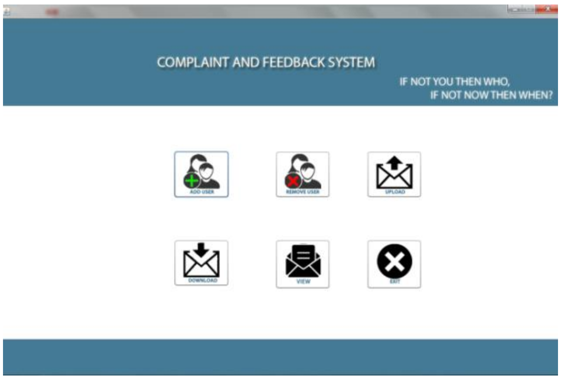


Figure 4:-HOD Dashboard

As shown in figure 3, the registered user will have to login first with valid credentials like username, password and designation. Designation includes HOD, staff and student. After successful login different user will be able to perform different operations as per designation.

HOD login and staff login will result in relatively same operations with only difference HOD will be able to add or register staff members and view their feedback too.

Student menu is shown in figure 5. The student will be able to upload and download complaints and submit feedback for college facilities as well as departmental faculties.

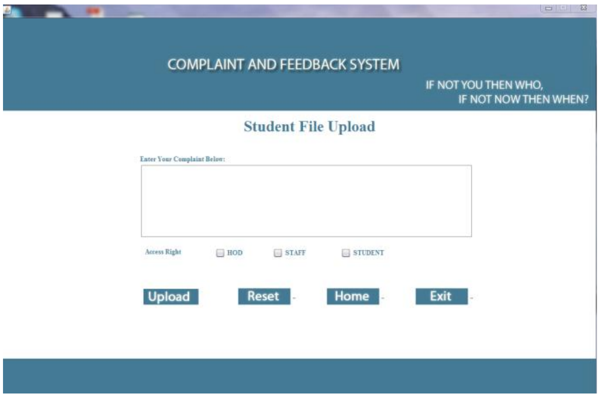
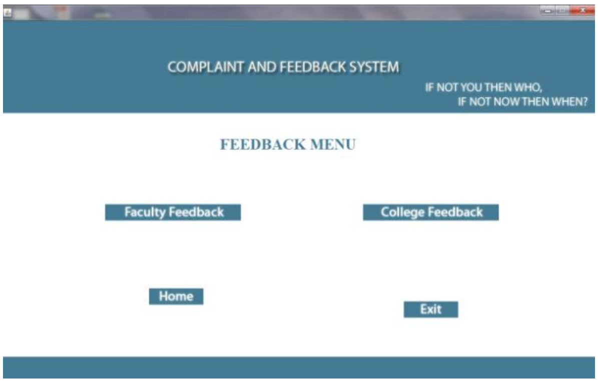


Figure 5:-Student Complaint Upload

While uploading complaint access rights are assigned i.e. who can view the compliant and messages is set.

DOI 10.17148/IJARCCE.2016.53117



COMPLAINT AND FEEDBACK SYSTEM

Figure 6:-Feedback Menu

Feedback menu is as shown in figure 6. Feedback submitted will view as the mean of all feedbacks. So feedback too will be anonymous.

5. CONCLUSION

Thus, authors propose an anonymous but secure authentication scheme for the data stored in cloud. User revocation is done and once a user is revocated cannot view the messages stored on the cloud. Also, authors propose text filtering to prevent improper and meaningless messages. This system can be useful for government and non government organizations.

Our aim is to promote paperless work. One can complain, give feedback and send notices on cloud storage.
In future, authors would like to provide document filtering and searchable encryption. Document filtering will improve the functionality of our system greatly. Searchable encryption will ease the searching of messages and other data.

REFERENCES

[1] A. Sahai and B. Waters, “Fuzzy Identity-Based Encryption,” Proc. Ann. Int’l Conf. Advances in Cryptology (EUROCRYPT), pp. 457-473, 2005.

[2] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data,” Proc. ACM Conf. Computer and Comm. Security, pp. 89-98, 2006.

[3] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-Policy Attribute-Based Encryption,” Proc. IEEE Symp. Security and Privacy, pp. 321-334, 2007.

[4] M. Chase, “Multi-Authority Attribute Based Encryption,” Proc. Fourth Conf. Theory of Cryptography (TCC), pp. 515-534, 2007.

[5] A.B. Lewko and B. Waters, “Decentralizing Attribute-Based Encryption,” Proc. Ann. Int’l Conf. Advances in Cryptology (EUROCRYPT), pp. 568-588, 2011.

[6] M. Green, S. Hohenberger, and B. Waters, “Outsourcing the Decryption of ABE Ciphertexts,” Proc. USENIX Security Symp., 2011.

[7] Sushmita Ruj, Member, Ieee, Milos Stojmenovic, Member, Ieee, And Amiya Nayak,” Decentralized Access Control With Anonymous Authentication Of Data Stored In Clouds” Ieee Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, February 2014.

[8] G. Wang, Q. Liu, and J.Wu, ”Hierarchical attribute - based encryption for fine-grained access control in cloud storage services," in Proceedings of the 17th ACM conference on Computer and communications security.

[9] Minu George, Dr. C.Suresh Gnanadhas, Saranya, "A Survey on Attribute Based Encryption Scheme in Cloud Computing ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.

[10] Jiangtao Ren, Sau Dan Lee, Xianlu Chen, Ben Kao, Reynold Cheng and David Cheung, "Naive Bayes Classification of Uncertain Data", 2009 Ninth IEEE International Conference on Data Mining.

[11] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah khan , "A Review of Machine Learning Algorithms for Text-Documents Classification", JOURNAL OF ADVANCES IN INFORMATION TECHNOLOGY, VOL. 1, NO. 1, FEBRUARY 2010.

DOI 10.17148/IJARCCE.2016.53117