Review of Attribute-based Keyword Search Authorization in Cloud

Mrs. Zabiha Khan
Assistant Professor, Dept. of CSE,
Ghousia College of Engineering, Ramanagaram,
Karnataka, India

Ms. Kamala Kumari, B. K, Rumana Iffath, Saima Ahmed, Zaiba Tabassum
UG Students
Ghousia College of Engineering,
Ramanagaram

***Abstract*— Mobile device can store or retrieve personal data from anywhere or any time with popularity of cloud computing. Cloud computing is one of the dominating infrastructure for enterprise as long as services are available to end user as there ocean of data are outsourced in the cloud as well. The data security problem in cloud becomes more and more sever and prevents for development of mobile cloud. Since mobile devices only have limited computing resources and power solutions, solution with low computational overhead are in great need for mobile cloud application. Also data in the cloud computing system lend to be out of control and privacy fragile. A mechanism to guaranty the ownership of data.**

We propose the attribute based encryption inspired methodology with first keyword search scheme with efficient user revocation (ABKS-UR) which enables suitable fine grained search authorization .ABKS-UR allows multiple owner to encrypt and outsource their data to cloud server independently. Designing a search result verification scheme will build resident of data user in the proposed service search system.

Index Terms*— Cloud computing, mobile cloud, fine grained owner enforced search authorization, attribute based keyword search, color based search.

I. INTRODUCTION

Cloud Computing provides us a means by which we can access the applications as utilities, over the internet. It allows us to create, configure, and customize the business applications online. Cloud Computing is an increasing mature model of enterprise IT infrastructure that provides on demand high quality applications and services from a shared pool of configuration computing resources. The cloud customers, individuals or enterprises, can outsource their local complex data system into the cloud to avoid the costs of building and maintaining a private storage infrastructure. The company or organization's private and sensitive information like personal files, company records, emails, etc which is to be shared among the selected company employees is stored and centralized into cloud server but mostly with an insecure feeling that anyone may hack these data that may be very risky for that company. Also the data owners and cloud server may not to be in the same trusted domain who put the outsourced unencrypted data, if any, at risk; the cloud server may leak data information to unauthorized entities or even be hacked. Because multiple cloud customers from the same or different organization can use the same resources or applications [1], certain security risks should be evaluated and solved before private and sensitive data, applications and system functionality are moved into the cloud in Fig.1.1. Multi-tenancy requires a policy enforcement mechanism, isolation, service levels, etc.

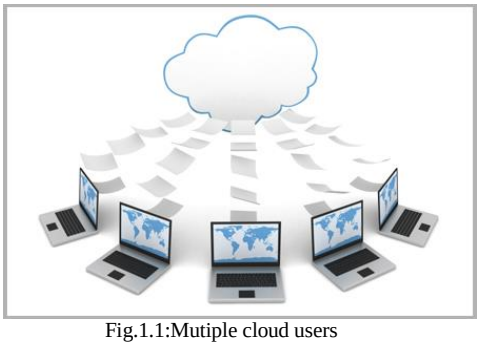

Fig.1.1:Mutiple cloud users

When Cloud computing, as a new paradigm of information technology, has been developed very quickly in recent years. The vast spread of Internet resources on the web and fast growth of service providers enabled cloud computing systems to become a large scaled IT service model for distributed network environments. Cloud computing is built on top of already existing Internet technologies and is delivered as a self- service utility. Three service models are: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [2].

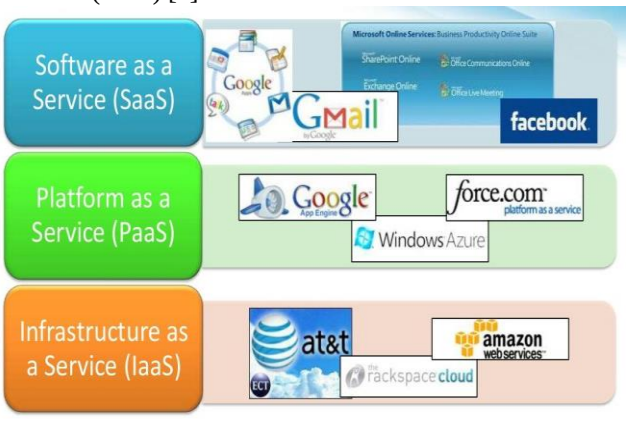

Fig.1.2:Service Models

Google, Microsoft Azure Platform, and Amazon Web Services are leading cloud computing vendors in the market of commercial system deployment. Regardless the utilized service model, cloud system can belong to one of the following cloud deployment models: Public, Community, Private or Hybrid[3].

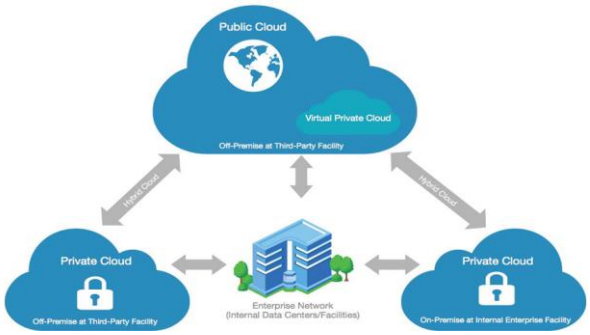

Fig 1.3:Deployment Model

The main characteristics of a cloud environment are abstraction and virtualization which make the technology to be perceived and applied completely in a different manner compared with existing traditional distributed systems. Cloud environment abstracts the implementation details of services and system from users and developers. Besides, resources in cloud computing systems become highly scalable through system virtualization which is achieved by means of resource pooling and sharing [2] [3]. Cloud computing has all the security issues associated with distributed applications on the Internet and plus other security . When sensitive data are outsourced to the cloud, data owners naturally become concerned with the privacy of their data in the cloud and beyond. Encryption-before-outsourcing has been regarded as a fundamental means of protecting user data.

This paper is organized as follows. In section II, we discuss about Keyword Guessing Attack.Then we discuss about Expressive Keyword Search in section III.Then Keyword based Search Scheme. Followed by Authentication Schemes for Session Passwords using Color and Image in section V.Followed by Attribute based encryption with keyword search scheme with user revocation (ABKS-UR) is discussed in VI. Finally conclusions in Section VII.

II. KEYWORD GUESSING ATTACK

To provide the privacy of the user who receive some computing services from the cloud, the users must encrypt their documents before outsourcing them to the cloud. Computation on outsourced encrypted data in the cloud rises some complexity to the system especially in the case when an entity would like to find some documents related to a special keyword. Searchable encryption is a tool for data owners to encrypt their data in a searchable manner. Generally, there exist two kinds of searchable encryption, namely symmetric (secret key) and asymmetric (public key) ones. But the public key searchable encryption schemes are vulnerable to the keyword guessing attack (KGA), an attribute-based keyword search scheme is secure against KGA.

In this the data users and the data owners are associated with a set of attributes and they receive their secret keys from a Trusted Authority (TA), corresponding to their attributes. So, the data owners encrypt their documents based on a search control policy and outsource the resulting ciphertext to the cloud server. In this model the data users are able to generate the required search token without any interactions with the data owners [1]. This kind of searchable encryption is called attribute-based keyword search scheme (ABKS) which is introduced by Zeng et al. Their scheme is presented with the inspiration of attribute-based encryption (ABE) [2] and the two proposed variants called ciphertext policy ABKS (CP-ABKS) and key policy ABKS (KP-ABKS). In KP-ABKS the cryptographic credentials are associated to the search control policy and in CP-ABKS the ciphertext of the keyword is associated to the search control policy [8].

The framework of our scheme involves four entities that is shown in Fig.2.1.

- Cloud Server (CS): It has strong computational capability and very large storage capacity. It stores and processes data and provides users with keyword search service.
- Data Owner (DO): This entity wants to share his data. For this propose he encrypts the desired data under certain access policy and finally outsources them to the CS.
- Data User(DU):This entity searches for certain keywords on the stored data in the CS. Using his secret key the DU generates search token and fuzzy search token and sends the latter to the CS for keyword searching. When the CS returns search results to the DU, the DU uses search token for keyword searching on receiving data.
- Trusted Authority (TA): The fully trusted third party manages a set of all attributes and distributes them between DOs and DUs. Also, TA generates secret keys for data users according to their attributes.

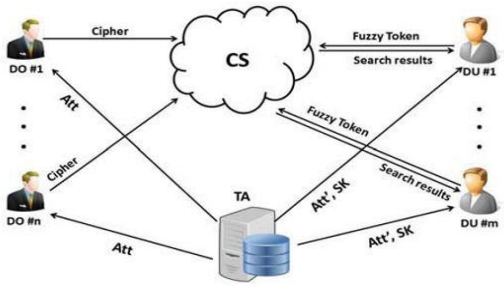

Fig.2.1. System model architecture.

We analyse security of the proposed scheme against KGA by assuming a probabilistic polynomial-time (PPT) adversary A, who may be an unauthorized DU. In this attack A has a valid search token and he knows the set of all keywords. He wants to find a keyword corresponding to the search token. The adversary runs the following algorithm for each keyword:

1) A encrypts the keyword and generates a keyword searchable ciphertext and then uploads the ciphertext to the cloud.

2) A sends the valid search token to the CS.

3) The CS sends search results to A. if the search results contains the ciphertext, A returns the keyword.

III. EXPRESSIVE KEYWORD SEARCH

The basic idea of this scheme is to modify a key-policy attributed-based encryption (KP-ABE) scheme constructed from bilinear pairing over prime-order groups. Without loss of generality, we will use the large universe KP-ABE scheme selectively secure in the standard model proposed by Rouselakis and Waters [3].In KP-ABE, a ciphertext is computed with respect to a set of attributes and an access policy is encoded into a user's private key. A ciphertext can be decrypted by a private key only if the set of attributes associated with the ciphertext satisfies the access policy associated with the private key. A KP-ABE scheme can be transformed to an expressive SE scheme by treating attributes as keywords to be searched, by directly transforming the key generation algorithm on attribute access structures to a trapdoor generation algorithm on keyword search predicates, and by using the decryption algorithm to test whether keywords in a ciphertext satisfy the predicate in a trapdoor. To keep our description compact and consistent, we will use access structure, policy and predicate interchangeably.

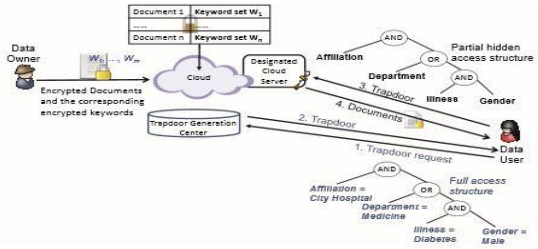

Fig.3.1. Architecture of expressive keyword search system

In order to hide keywords in a ciphertext, inspired by the "linear splitting" technique in [4], we firstly split ciphertext components corresponding to every keyword into two randomized complementary components. Thus, even though the ciphertext still contains information about the keywords, this information is computationally infeasible to obtain from the public parameter and the ciphertext. We secondly re-randomize trapdoor components corresponding to every keyword associated with an access structure to match the spited components in the ciphertext. The architecture of the keyword search system is shown in Fig.3.1.

IV. KEYWORD-BASED SEARCH SCHEME

Normally, keyword search over encrypted data schemes share the same search model. In this model, we need to pay attention to the cloud server. The whole process of keyword search goes as following statement. First, the data owner need to construct index for the document set, uses an encryption way to encrypt the document collection and the index, and then outsources them into the cloud server.

A. Keyword search scheme based on text document->

Keyword search can be divided into accurate keyword search and fuzzy keyword search. In accurate keyword search range, if the query meets the keyword set, it will return the corresponding document.

B. Keyword search scheme based on XML document->

The keyword search schemes based on XML document with structural information are also proposed in recent years present a solution for efficient evaluation of tree pattern queries (TPQs) on encrypted XML documents. Rao et al. [5] introduce a new approach multiple attribute group decision making (MAGDM) for dealing with fuzziness in ranking and selection of alternatives with respect to multiple attributes. S. Selvaganesan et al. [6] put forward a new keyword search approach named XML keyword search dual indexing and mutual summation algorithm (XDMA) to solve keyword ambiguity and query result grading problem. Wang et al. [7] show that NP-hard is a secure and optimal encryption. They propose to keep metadata consisting of both structure and value indices on server for speeding up query process.

C. Novel color-based keyword search scheme for encrypted office document->

In color-attribute-based multi-keyword search over encrypted cloud OOXML data, we use two vectors on behalf of the keywords and the corresponding color. Given a query, we divide it into two vectors and separately matching operations. As a result, the scheme returns a top-k set as correct as possible meeting user's requirements

Our scheme to support keyword search based on color is as follows.

1. Setup: In the initialization stage, the data owner generates a symmetric key SK, including: (a) a randomly generated vector with n-bit; (b) two m*m randomly invertible matrices {M1, M2}. So SK is combined by{S, M1, and M2}.
2. GenIndex: Firstly, we extract keyword set together with color attribute from the document collection. An unencrypted index tree is built according to buildIndexTree (). Then, index vector with TF weight and color attribute vector are stored in node of tree. After that, the data owner generates two random vectors to encrypt the tree.
3. GenQuery: For the query keywords, two n-dimension query vectors Q and Q' are generated presenting IDF weight and the color attribute of corresponding keyword, separately. Especially, the color attribute is composed of the black and the red color.0 presents black and 1 presents red.
4. Search: Given a query, we need to match the color vector first, once failed to be matched, it means keyword with color does not exist in the documents. And then calculate the product of query and keyword vector, we pick up the bigger result of the product and go through the subtree. Once matched, we priory pick the sub-tree and calculate the product. At last, after traversing the tree, it will

return the fit documents which may meet users' requirement.

V. AUTHENTICATION SCHEMES FOR SESSION PASSWORDS USING COLOR AND IMAGE

Authentication technique consists of 3 phases: registration phase, login phase and verification phase. During registration, user enters his password in first method or rates the colors in the second method. During login phase, the user has to enter the password based on the interface displayed on the screen. The system verifies the password entered by comparing with content of the password generated during registration [12].

A. Pair-based Authentication scheme:

During registration user submits his password. Minimum length of the password is 8 and it can be called as secret pass. The secret pass should contain even number of characters. Session passwords are generated based on this secret pass. During the login phase, when the user enters his username an interface consisting of a grid is displayed. The grid is of size 6 x 6 and it consists of alphabets and numbers. These are randomly placed on the grid and the interface changes every time.

The first letter in the pair is used to select the row and the second letter is used to select the column. The intersection letter is part of the session password. This is repeated for all pairs of secret pass. Fig.4.1 shows that L is the intersection symbol for the pair "AN". The password entered by the user is verified by the server to authenticate the user. If the password is correct, the user is allowed to enter in to the system. The grid size can be increased to include special characters in the password.

Fig.4.1: Intersection letter for the pair of AN

B. Hybrid Textual Authentication Scheme

During registration, user should rate colors as shown in figure.4.2. The User should rate colors from 1 to 8 and he can remember it as "RLYOBGIP". Same rating can be given to different colors. During the login phase, when the user enters his username an interface is displayed based on the colors selected by the user. The login interface consists of grid of size 8×8. This grid contains digits 1-8 placed randomly in grid cells. The interface also contains strips of colors as shown in figure.4.3. The color grid consists of 4 pairs of colors. Each pair of color represents the row and the column of the grid.

Fig.4.2: Rating of colors by user.

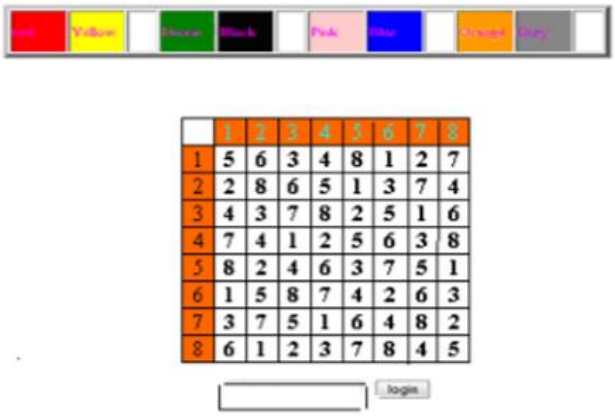

Fig.4.3: Login interface.

Figure.4.3 shows the login interface having the color grid and number grid of 8 x 8 having numbers 1 to 8 randomly placed in the grid. Depending on the ratings given to colors, we get the session password. As discussed above, the first color of every pair in color grid represents row and second represents column of the number grid. The number in the intersection of the row and column of the grid is part of the session password. Consider the figure 9 ratings and figure 10 login interface for demonstration. The first pair has red and yellow colors. The red color rating is 1 and yellow color rating is 3. So the first letter of session password is 1st row and 3rd column intersecting element i.e,3. The same method is followed for other pairs of colors. For figure 10 the password is "3573". Instead of digits, alphabets can be used. For every login, both the number grid and the color grid get randomizes so the session password changes for every session.

C. Security Analysis

As the interface changes every time, the session password changes. This technique is resistant to shoulder surfing. Due to dynamic passwords, dictionary attack is not applicable. Hidden camera attacks are not applicable to PDAs because it is difficult to capture the interface in the PDAs.

- Dictionary Attack: These are attacks directed towards textual passwords. Here in this attack, hacker uses the set of dictionary words and authenticate by trying one word after one. The Dictionary attacks fails towards our authentication systems because session passwords are used for every login.
- Shoulder Surfing: These techniques are Shoulder Surfing Resistant. In Pair based scheme, resistance is provided by the fact that secret pass created during

registration phase remains hidden so the session password can't be enough to find secret pass in one session. In hybrid textual scheme, the randomized colors hide the password. In this scheme, the ratings decide the session password. But with session password you can't find the ratings of colors. Even by knowing session password, the complexity is 8 4. So these are resistant to shoulder surfing. Guessing: Guessing can't be a threat to the pair based because it is hard to guess secret pass and it is 36 4. The hybrid textual scheme is dependent on user selection of the colors and the ratings.

International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.3, May 2011

If the general order is followed for the colors by the user, then there is a possibility of breaking the system. Brute force attack: These techniques are particularly resistant to brute force due to use of the session passwords. The use of these will take out the traditional brute force attack out of the possibility.

- Complexity: The Complexity for Pair-Based Authentication Scheme is to be carried over the secret pass. For a secret pass of length 8, the complexity is 368. In the case of the Hybrid Textual Authentication Scheme the complexity depends on colors and ratings. The complexity is 8! if ratings are unique, otherwise it is 8^8.

VI. ATTRIBUTE BASED ENCRYPTION WITH KEYWORD SEARCH SCHEME WITH USER REVOCATION (ABKS-UR)

There has been a great interest in developing attribute based encryption [28], [29], [30], [31] due to its fine grained access control property. Goyal et al. [28] designed the first key policy attribute-based encryption (KP-ABE) scheme, where ciphertext can be decrypted only if the attributes that are used for encryption satisfy the access structure on the user private key. Under the reverse situation, CP-ABE allows user private key to be associated with a set of attributes and ciphertext associated with an access structure. CP-ABE is a preferred choice when designing an access control mechanism in a broadcast environment.

The system framework of our proposed ABKS-UR scheme involves three entities: *cloud server*, many *data owners*, and many *data users*, as shown in Fig.5.1.

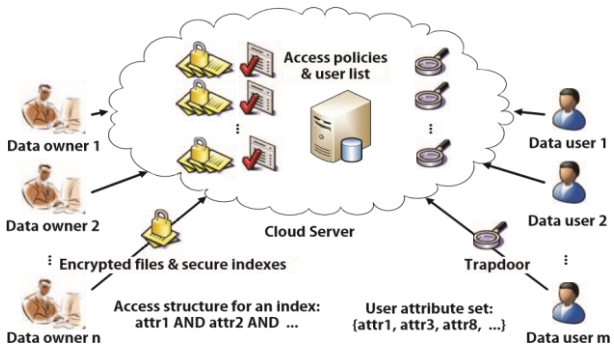

Fig.5.1: Framework of authorized keyword search over encrypted cloud data.

In addition, a trusted authority is implicitly assumed to be in charge of generating and distributing public keys, private keys and re-encryption keys. To enforce fine-grained authorized keyword search, the data owner generates the secure indexes with attribute-based access policies before outsourcing them along with the encrypted data into the CS. Note that we can encrypt data by any secure encryption technique, such as AES, which is outside the scope of this paper. To search the datasets contributed from various data owners, a data user generates a trapdoor of keyword of interest using his private key and submits it to the CS. So as to accelerate the entire search process, we first enforce the coarse-grained *dataset search authorization* with the *per-dataset* user list such that search does not need to go to a particular dataset if the user is not on the corresponding user list. Next, the finegrained *file-level* search authorization is applied on the authorized dataset in the sense that only users, who are granted to access a particular file, can search this file for the intended keyword. More precisely, the data owner defines an access policy for each uploaded file. The CS will search the corresponding datasets and return the valid search result to the user if and only if the attributes of the user on the trapdoor satisfy the access policies of the secure indexes of the returned files, and the intended keyword is found in these files. Our proposed ABKS-UR scheme in the cloud aims to achieve the following functions and security goals:

1) Authorized Keyword Search.
2) Supporting Multiple Data Contributors and Data Users.
3) Efficient User Revocation
4) Authenticity of Search Result.

VII. CONCLUSION

In this paper, we design the verifiable attribute based keyword search scheme in the cloud environment, which enables scalable and fine-grained owner-enforced encrypted data search supporting multiple data owners and data users. In the proposed scheme keyword guessing attack (KGA), the data owners generate two fuzzy and exact searchable ciphertexts according to a search control policy and send them to the cloud. We have shown that in this model, our proposed scheme is secure against KGA. In order to allow a cloud server to search on encrypted data without learning the underlying plaintexts in the public key setting, Boneh proposed a cryptographic primitive called public-key encryption with keyword search (PEKS).We focused on the design and analysis of public-key searchable encryption systems in the prime-order groups that can be used to search multiple keywords in expressive searching formulas. We present the framework of the searchable encryption, and then descript the existing keyword-based search technologies. We use two vectors on behalf of the keywords and the corresponding color. Given a query, we divide it into two vectors and separately matching operations. As a result, the scheme returns a top-k set as correct as possible meeting user's requirements. In this paper we also brief out about the authentication of session password using color and image scheme.

REFERENCES

[1] Q. Zheng, S. Xu, and G. Ateniese, "Vabks: Verifiable attribute-based keyword search over outsourced encrypted data," in INFOCOM, 2014 Proceedings IEEE. IEEE, 2014, pp. 522–530.

[2] Vahid Yousefipoor, Mohammad Hassan Ameri, Javad Mohajeri, "A Secure Attribute Based Keyword Search Scheme against Keyword Guessing Attack", 2016 8th International Symposium on Telecommunications (IST'2016).

[3] Y. Rouselakis and B. Waters, "Practical constructions and new proof methods for large universe attribute-based encryption," in 2013 ACM SIGSAC Conference on Computer and Communications Security,CCS'13,Berlin,Germany,November4-8,2013. ACM, 2013, pp. 463–474.

[4] Hui Cui, Zhiguo Wan, Robert H. Deng, Guilin Wang, and Yingjiu Li, "Efficient and Expressive Keyword Search Over Encrypted Data in Cloud", 1545-5971 (c) 2016 IEEE.

[5] Rao, D. Rajeswara, and Lavanya Susanna. Efficiently Retrieving top k Search Results over XML Data. International Journal of Research in Computer Engineering & Electronics 3, 2014.

[6] Selvaganesan, S. , Su-Cheng Haw, and Lay-Ki Soon. Effective XML Keyword Search Using Dual Indexing Technique, Information Technology Journal, 2014, 13(4): 643-651.

[7] Wang H, Lakshmanan L. Efficient secure query evaluation over encrypted XML databases, Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment, pp. 127138, 2006.

[8] Zhangjie Fu, Member, IEEE, Jie Xi, Jin Wang, Xingming Sun, "Document Attribute-based Keyword Search over Encrypted Data", 978-1-4799-5390-5/14 $31.00 © 2014 IEEE.

[9] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting Your Right: Attribute-based Keyword Search with Fine-grained Owner enforced Search Authorization in the Cloud," in IEEE INFOCOM, pp. 226-234, 2014.

[10] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," IEEE TPDS, vol. 24, no. 1, pp. 131143, 2013

[11] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized private keyword search over encrypted data in cloud computing," in Proc. of IEEE ICDCS, pp. 383-392, 2011.

[12] M Sreelatha 1, M Shashi 2, M Anirudh 1, Md Sultan Ahamer 1, V Manoj Kumar, Authentication Schemes for Session Passwords using Color and Images International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.3, May 2011

[13] Z. Zheng, X. Liu, L. Yin, Z. Liu "A Hybrid password authentication scheme based on shape and text" Journal of Computers, vol.5, no.5 May 2010.

[14] H. Zhao and X. Li, "S3PAS: A Scalable Shoulder-Surfing Resistant Textual-Graphical Password Authentication Scheme," in 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW 07), vol. 2. Canada, 2007, pp. 467-472.

[15] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in Proc. of IEEE INFOCOM, pp. 1-9, 2010.