Dynamic Data Mining Tool on Distributed Encrypted Data

[1] Pranali M. Sonawane , [2] Vikas K. Kumawat , [3] Vijayalaxmi L. Balsane , [4] Prachi B. Suryavanshi , [5] Ramnath Banerjee

[1,2,3,4,5] Department of Computer Engineering, Pune University, Dr. D.Y. Patil Institute of Engineering & Technology
Ambi, Pune, Maharashtra, India

Abstract- Now-a-days the security of the system matters because system gets easily hacked, so deal with this problem. Here we introduce the 3d key generation. It generated from RGB colors and their priority. User data distributed in smaller parts and uploaded server side database. To maintain the security of data, data encryption and decryption is introduced and this task is done by AES and RC6 algorithm. J48, Fuzzy c-means clustering algorithm, apriori classification algorithm are online data mining. Using online data mining tool we can predict multiple result with graphical representation.

Keywords- *Data distribution theory, Data mining, security, Data encryption and decryption.*

1. Introduction

The web application provides users a decision model which provides a better security by distributing data over multiple web service providers in such a way that, none of the SP can successfully retrieve meaningful information from the data pieces allocated at their servers. Using online data mining tool we provide the user with better assurance of availability of data, by maintaining redundancy in data distribution. In this case, if a service provider goes out of service or corrupted, then user still can access his data by retrieving it from other service providers. From the business point of view, since web data storage is a subscription service, the higher cost will be paid by user only when the data redundancy is higher. Thus, we provide an Optimization scheme to handle the tradeoff between the costs that a web application user is willing to pay to achieve a particular level of security for his data.

In other Words, we provide a scheme to enlarge the security for a given budget for the web data. In addition to that, Privacy preservation and data integrity are two of the most critical security issues related to user data. In conventional paradigm, the organizations had the physical control of their data and hence have an ease of implementing better data security policies. But in case of web application, the data is stored on an autonomous business party that provides data storage as a subscription service. The users have to trust the web service provider (SP) with security of their data. It has been noted that the critical conditions of the privacy issues in web application, and pointed out that obtaining information from a third party is much easier than from the creator himself. Following the design of paradigm shift, the security policies also evolved from the conventional cryptographic schemes applied in centralized and distributed data storage, for enabling the data privacy.

2. Literature Survey

In web application user's data is stored to a solo service provider. It is not secured according to user view. So in this paper, we propose a multi-web storage model in web application which holds an economical distribution of data among the available SPs, to provide customers with data availability as well as secure storage.

In Wireless Sensor Networks (WSNs) data security plays an important role where confidentiality, authentication, integrity are given importance. This paper, propose a User Authentication scheme for WSNs, which employs RGB color cube algorithm and Armstrong number for data security.

Transferring data over a network is not secure and hacker can hack or can change the data in between communication. To deal with this problem data encryption & decryption is done using AES and RC6 algorithm. Senders send the data which is encrypted and added with some keys. At receiver side data is decrypted using that key.

Fuzzy c-means algorithm with spatial constraint (FCM_S) is used for Data Mining, but it has two major disadvantages. 1) Data get noised and didn't get the knowledge of it; 2) Segment time is depends on data size. In this paper, too overcome this disadvantages Fuzzy c-means clustering algorithm is used.

3. Proposed System

The system aim to incorporate the previous system advantages and extends to find the unauthorized user, to prevent the unauthorized data access for preserving data integrity and use of data mining.

3.1 Session Authentication

For Session authentication, 3D key generation is introduced. It is basically a key generated from RGB color cube algorithm and Armstrong number. Different steps of this algorithm is as followed-

1. RGB Based Authentication
2. RGB Color Cube
3. User registration
4. User authentication on Login

Every time key generated randomly when user tries to login and it is sent to his/her email-id. After conforming 3D key he/she got login.

3.2 Data Uploading and Distribution

Data uploading on this web application is dynamic in nature. User can store any kind of data like file, Database etc. System asks for the priorities of data and distributes data into high level and low level entities. This data then stores into number of service provider. It help user to secure his/her data from unauthorized access.

3.3 Data Encryption Using RC6 & AES

Securing data on web is one of the major issues while storing and retrieving data. AES algorithm and RC6 is used for encrypting and decrypting the data. Advanced Standard Encryption (AES) is used for file like .xls, Access file etc. and RC6 is used for databases (MySQL, SQL server etc.).

3.3.1 AES Algorithm for Encryption

The Advanced Encryption Standard (AES) is a symmetric-key block cipher. AES allows for three different key lengths: 128, 192, or 256 bits

AES involve main 4 steps:-
1. Substitution
2. Shifting of rows
3. Mixing of columns
4. Adding round key

AES Encryption consists of 10 rounds of processing for 128-bit keys. Each round of processing includes one single-byte based substitution step, a row-wise linear step, a column- wise mixing step, and the addition of the round key. The first transformation of Sub Bytes is used at the encryption site. To substitute a byte, we involved the byte as two hexadecimal digits. In shifting of rows the 1st row is not to be shifted whereas 2nd 3rd and 4th row are shifted by one, two and three bytes respectively. The Mix Columns transformation operates at the column level; it transforms each column of the state to a new column. AddRoundKey precedes one column at a time. AddRoundKey combine a round key word with every state column matrix; the operation in AddRoundKey is matrix addition.

3.3.2 RC6 Algorithm

RC6 is next version of RC5. It is specified as RC6-w/r/b where w is word size, r is number of rounds for encryption or decryption and b is length of encryption key in bytes. RC6 work with four registers A, B, C, D and takes plaintext as input, output is cipher text at end of encryption. First byte of plaintext or ciphertext placed to A, and last byte is placed to D. (A,B,C,D)=(B,C,D,A) is used which mean parallel assignment of values on right registers on the left. In the end it will result as chipertext for encryption or plaintext for decryption.

3.4 Data Mining

Fuzzy c-means clustering algorithm, Apriori algorithm and J48 classification algorithm is used for data mining. Classification tree is generated as the result of data mining. Fuzzy c-means clustering algorithm is run on client side. Apriori algorithm and J48 classification algorithm are online software in this web application.

3.4.1 Fuzzy c-means clustering Algorithm

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. In No of ‘k’ clusters, it will find the nearest point between these clusters according to classification called ‘K-Center’. As per algorithm the center is the average of all the points in cluster. By using means of these clusters it will find new centroids. Steps are repeated till the result is not found.

3.4.2 Apriori Algorithm

Aprioriis algorithm for frequent item set mining and association rule learning over transactional databases. The purpose of Apriori algorithm is to find association between the sets of data. It is some time to refer as "Market Basket Analysis". Each set of data has number of items which is called as transaction. Apriori result is a set of rules that tell us how often items are contained in sets of data.

3.4.3 J48 classification Algorithm

J48 Algorithm is involved with following steps:
1. Classification Tree 2. Split to consideration
3. Tree construction 4. Over fitting and Pruning

Classification of data can be based on the splitting of data. Withdrawing the impurities of splinted data could be done by split to consideration. Tree construction starts at the top of the tree with all the data. Complexity and error reduced by over fitting and pruning.

4. Tables, Figures and Equations

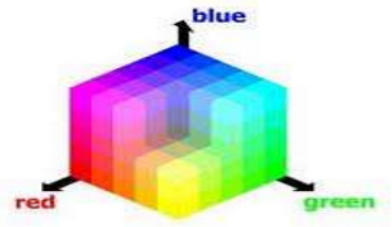

Figure 1: Color Cube

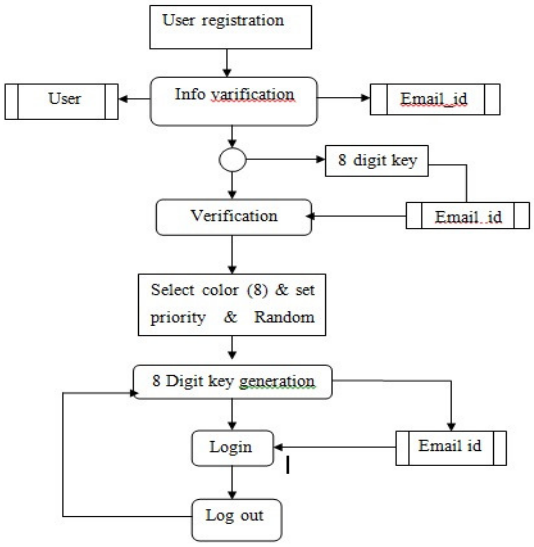

Fig 2: 3D Generation

5. Result

User is successfully registered to our system. At login time 3D key is sent to his/her email- id. After entering 3D key he/she get logged in to system. User database successfully got uploaded to server. He/she trying to fetches the data and desired output had come.

6. Conclusion

The proposed model explains a Secured Cost-Effective Multi-Web Storage (SCMCS) in web application, which seeks to provide each customer with better decision making for data storage.. Using SP user budget for storing data and quality of services can be achieved. By using the AES algorithm for encryption we provide security to data. Fuzzy c-means algorithm is used for data classification. By dividing and distributing customer's data, proposed model shows its ability of providing a customer with a secured storage under his affordable budget.

Acknowledgments

We are grateful to Prof. Ramnath Banerjee Sir and Prof. Sharmila Chopade madam for their valuable advice and guidance during the preparation of this project.

References

[1] C. Wang, Sherman S.-M. Chow, Q. Wang, K. Ren, W. Lou, "Privacy preserving public auditing for secure web storage", in InfoCom2010, IEEE, March 2010.

[2] B. Adler, Load balancing in the web: Tools, tips and techniques, http://www.rightscale.com/infocenter/whitepapers/ Load-Balancing-in-the-Web.pdf, 2012P.

[3] Ismail Butun and Ravi Sankar,2011." Advanced Two Tier User Authentication Scheme for Heterogeneous Wireless Sensor Networks". 2nd IEEE CCNC Research Student Workshop.

[4] Akash Kumar Mandal, Chandra Parakash, Mrs Archana Tiwari "Performance Evaluation of Cryptographic Algorithms: DES andAES", IEEE Trans. on Electrical, Electronics and Computer Science, 2012.

[5] G. C. Karmakar and L. S. Dooley, A generic fuzzy rule based image segmentation algorithm, Pattern Recognition Letters, 23(10) (2002) 1215- 1227.

[6] ShufenZhang,ShuaiZhang,Xuebinchen, ShangzhuoWu, "Analysis and Research of Web application system Instance", 2010 second international Conference on Future Networks, IEEE.

[7] C.S. Yeo, R. Buyya1, M.D. de Assunção, et al. Utility Computing on Global Grids. Technical Report, GRIDS-TR-2006-7, Grid Computing and Distributed Systems Laboratory, The University of Melbourne, Australia, 2006.

[8] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring data storage security in web application," in Proc. of IWQoS'09, July 2009, pp.1–9.

Pranali M. Sonawane
Position: Student
Degree: Diploma in Computer Engineering (2011)
Bachelor in Computer Engineering (2014)

Vikas K. Kumawat
Position: Student
Degree: Bachelor in Computer Engineering (2014)

Vijayalaxmi L. Balsane
Position: Student
Degree: Diploma in Computer Engineering (2011)
Bachelor in Computer Engineering (2014)

Prachi B. Suryavanshi
Position: Student
Degree: Diploma in Computer Engineering (2011)
Bachelor in Computer Engineering (2014)

Prof. Ramnath Banerjee
Position: Professor
Department of Computer
Dr. D.Y. Patil Institute of Engineering & Technology
Ambi, Pune.