International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Novel Approach to Improve the Privacy of Information Brokering in Semantic Web

Supriya S. Sankpal[1], Rupali A. Mahajan[2]

[1]ME (COMP) Student, BSIOTR (W), Wagholi, Pune, Maharashtra, India

[2]Professor, BSIOTR (W), Wagholi, Pune, Maharashtra, India

Abstract: ***As enormous structured, semi-structured and unstructured data are collected and archived by organizations in many realms ranging from business to health networks to government agencies, the needs for efficient yet secure inter-organization information sharing via on-demand access naturally arise. Now a day, information brokering system has honest assumptions on brokers who can fulfill the requirements of user by locating right data provider where required data is present. Moreover, with increasing concerns on protecting the sensitive and/or proprietary data, the organizations prefer sharing data in a more secure and privacy-preserving manner, instead of establishing a purely full trust relationship on brokers and releasing the control over the shared data. In this paper, we propose a new privacy-preserving information brokering system (PPIBS), with privacy enhancing encryption algorithms such as AES for selective encryption which protect the identity of the data requestor who forward the request to broker and two tier encryption using Vigenere cipher and Selective Reverse Circle Cipher and Huffman algorithm to enrich the data privacy.***

Keywords: Access Control, Encryption, Information Brokering System, Privacy, Secrete Key

1. Introduction

Today's organizations often operate across organizational boundaries. They raise strong needs for efficient and secure information sharing to facilitate extensive collaborations. Organizations (e.g. enterprise, government agencies, and libraries, “Smart” Home) may consist of data requestor, data provider or both. Information sharing is done in different ways in different applications. However, early approaches on information sharing do not satisfying new requirements of these inter-organizational collaborations. Most of the Information brokering system uses either the request–reply model to build client-server connections depending upon on-demand information access where client-server peers are absolutely autonomous or the distributed database model, where all client-server peers with insufficient autonomy. These two models are not suitable for many new applications, like healthcare or law enforcement system, in which information sharing within organization occurs in controlled manner because of some legal reasons.

To better understand such requirements, we overview the unique needs of such inter-organization collaboration by considering an example in the health-care domain. Large-scale health information infrastructures, such as Regional

Health Information Organization (RHIO), are being developed to share medical information (e.g., patient records) collected by collaborative health providers (e.g., hospitals). Each health provider is authorized by its patients to collect medical information. Since the data is private and sensitive, the health providers are responsible for not leaking patient records to irrelevant parties. The health providers desire to share their data to fulfill collaboration, however, they prefer to do it in a restricted and controlled fashion. Data requestors, such as doctors, need to be able to retrieve the medical records with precision and not be distracted by “noisy" data [1].

In such situations, sharing an entire copy of the information with other data providers or collecting data in centralized manner become unrealistic. To tackle this situation federated database system [2] has been planned that are autonomous, heterogeneous, to handle local data and provide unified data access. However centralized data still introduces data heterogeneity, privacy and trust issues.

As the data provider consist of sensitive and autonomous data, a more efficient solution is to create a data centric overlay consisting of data sources and set of third party i.e. brokers which routes the user’s queries to specific data source according to contents of query. Information brokering system is a system where on-demand data is accessed through set of brokers [3].While the Information brokering system provides flexibility and severs autonomy, privacy concern of sensitive information arises, as brokers as no longer assumed honest. The broker may provide the data to the third party and thus vulnerable to be corrupted by insiders or by outsiders

To tackle these challenges, we present a new approach for preserving the privacy of Information Brokering System in Semantic Web, which provides a solution to enhance privacy of data requestor and data forwarded from data provider. The rest of the paper is organized as follows, Section II literature survey for the Information Brokering System. The implementation design details are provided at section III, Experimental result is illustrated in Section IV. In Section V a conclusion and future scope is mentioned.

2. Literature Survey

In [4] a novel approach that combines cryptography with authorizations, thus enforcing access control via selective encryption. The paper presents a formal model for access control management and illustrates how an authorization policy can be translated into an equivalent encryption policy while minimizing the amount of keys and cryptographic

tokens to be managed. The paper also introduces a two-layer encryption approach that allows the data owner to outsource, besides the data, the complete management of the authorization policy itself, thus providing efficiency and scalability in dealing with policy updates.

In [5] advanced encryption standard (AES) have been implemented using MATLAB software. The primary goal of this paper is to improve level of security. The implemented encryption technique is analyzed by using a parameter called Avalanche effect. Plaintext and encryption key are mapped in binary code before encryption process. Avalanche Effect is calculated by changing one bit in plaintext keeping the key constant and by changing one bit in encryption key keeping the key constant, Experimental results shows that the proposed algorithm exhibit significant high Avalanche Effect which improves the level of the security.

In [6] defines access control policies depends on data attributes to obtain properties scalability, data confidentiality and fine-grandness concurrently in the systems. To achieve goal it combines techniques like attribute-based encryption (ABE), proxy re-encryption, and lazy re-encryption. This scheme can enable the data owner to delegate most of computation overhead to powerful cloud server. Confidentiality of user access privilege and user secret key accountability can be achieved.

In [7] to fulfill the requirements of user, approach of balanced access control system is presented. Here, all the users of system must obey the policies of the system defined administrator. Policies consist of private keys of system users and access structure over attributes (resources) they can access. The users have freedom for designing their own access control structure for different files. A user has control on whom and under what circumstances can access their documents. This approach, a system administrator achieves users' privacy and system's security by providing the access control policies on documents. Mainly this system is depends on two techniques of attribute-based encryption: KP-ABE and CP-ABE. Firstly, access policies are placed into decryption keys, and then bind it with cipher texts.

In [8] depending on secrete sharing an approach of key-assignment is developed. By using this approach key is derived from two different techniques, and then combined them as a single technique. The number of public tokens necessary for the previous over-encryption scheme and their scheme are compared and it illustrate that this approach of sharing is more efficient than previous one.

In [10] Personal Health Record is a model a health information exchange, where patient data is outsourced to the third party so the privacy issue arises. To assure the patients' control over access to their own PHRs, it is a promising method to encrypt the PHRs before outsourcing. In this, proposes a novel patient-centric framework and a suite of mechanisms for data access control to PHRs stored in semi-trusted servers. To achieve fine-grained and scalable data access control for PHRs, we leverage attribute based encryption (ABE) techniques to encrypt each patient's PHR file. It also enables dynamic modification of access policies or file attributes, supports efficient on-demand user/attribute revocation and break-glass access under emergency scenarios.

In [12] Reverse circle cipher is symmetric polyalphabetic block cipher uses a concept called circular substitution and reversal transposition. It combines the simple character level displacement principle of the Caesar cipher, the distribution principle of the Vernam polyalphabetic cipher and the diffusion principle of the transposition cipher [10]. The system provides security even with white box and grey box models in addition to black box models of attacks [7][8].

3. System Overview

Block diagram0

The Figure 1 is proposed system architecture of Privacy-Preserving Information Brokering System having four steps:

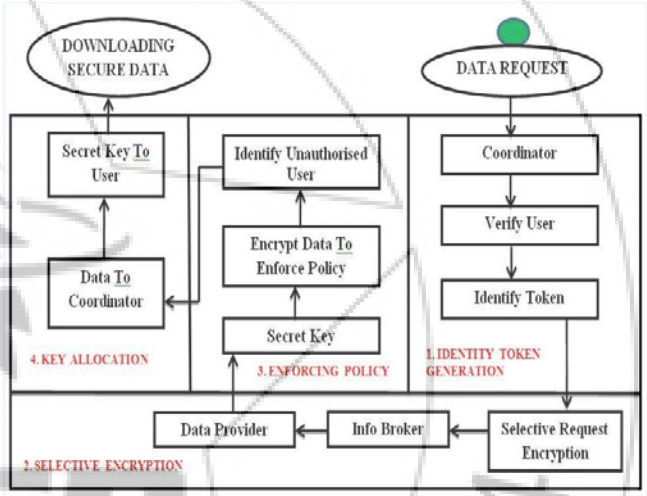

Figure 1: System Architecture

- **Implementation Steps:**

1: Identity token generation:
The data requestor request for data by entering many numbers of attributes which can contain name, address, gender and much other personal information. This request actually goes to a coordinator system where an identity token will be send back to data requestor as an acknowledgment.

$$(a)\ f(x) = \sum_{i=0}^{i=n} \text{Ui}$$

$$(b)\ Ps = P\ (f(X))$$

Where,
f(x) : function to concatenate all string of the field from the user profile.
Ui : each profile attribute.
Ps : Pseudo identity.
P(f(x) : Random Function to calculate Pseudo identity token.
Algorithm for Identity token generation:

Input : Set U ={ u1,u2,u3……un}
Output: pseudonym (Ps)

1. Start
2. Get the User Profile attribute set U
3. Convert all the attributes to String type
4. Concatenate all the String to get a single
5. String
6. Get the auto incremented User ID as 1
7. x=ID mod 7
8. for i=0 to String length
9. Fetch x^{th} character from the String
10. Continue till 7 characters are selected
11. Concatenate all the 7 characters
12. Return key
13. stop

2: Selective Encryption:
For preserving the privacy of the data requestor some of the confidential data such as Name, Address, and Hospital and doctor name of the requestor will be encrypt using AES algorithm and forward to the broker to get required data from data provider.

3: Enforcing policy:
In this module of our proposed system, the brokers forward the selected data which are already got from coordinator to the data provider with a identity token. The data provider will encrypt the data using two tier encryption algorithms:
1. Vigenere cipher algorithm.
2. Selective Reverse circle cipher algorithm with symmetric key.

First tier Encryption

$$C_i = T_i + K_i \ (mod\ m)$$

Where
C_i : i^{th} character of the ciphertext
T_i : i^{th} character of the open text
K_i : i^{th} character of the key phrase (if the key phrase is shorter than the open text, which is usual, then the key phrase is repeated to match the length of the open text
m : Length of the alphabet

Algorithm for Vigenere Cipher encryption

Input :Plaintext file
Output: Ciphertect file

1. Start
2. For each character Ci of the plain text
3. Get the ith character of the key
4. Repeat key to match the length of the open text
5. get length of the alphabet m
6. M=mod m
7. C_i=K_i(character of key) +M
8. End of for
9. Stop
10.

Second Tier Encryption
For providing double security the data is encrypt again using Selective Reverse Circle encryption algorithm for text based files.

$$Pi = f^{-1}(Ci, k(0 + len(k)i) \ldots\ldots\ldots\ldots\ldots (d)$$

$$R_{cp}=\sum_{i=0}^{R} \sum_{j=0}^{B} Pi \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (e)$$

Algorithm for Reverse Circle Cipher Encryption

Input : Encrypted text file with Vigenere Cipher
Output : Ciphertext file

1. Start
2. Get Input String S
3. Initialize a String ENC as empty
4. Divide the string S in N blocks of size 10 characters
5. **for** I =1 to N
6. Let String BS =10 character of each block
7. rotate block with I characters in **clock wise**
8. **for** j=1 to 10
9. substitute each character
10. Replace character
11. **End of inner for**
12. ENC=ENC+BS
13. **End of Outer for**
14. Stop

If the data file provided by data provider is word or PDF file then that will be encrypted by using Huffman Compression algorithm.

Algorithm for Huffman Compression

Input : Word or PDF file
Output :Encrypted file

1. Start
2. Create a leaf node for each symbol and add it to the priority queue **Q**.
3. While there is more than one node in the queue **Q**
4. Remove the two nodes of highest priority (lowest probability) from the queue
5. Create a new internal node with these two nodes as children and with probability equal to the sum of the two nodes' probabilities.
6. Add the new node to the queue **Q**.
7. **End of While**
8. The remaining node is the root node and the tree is complete.
9. Stop

4. Key Allocation

In this final step, a special key i.e. identity token will get generated form coordinator for the end user access control to get the original data of data provider.

5. Conclusion and Future Scope

The existing Information Brokering System is having privacy leakage of data requestor privacy, data privacy and metadata privacy. In this paper we introduced a new approach for enriching the privacy of data shared within Information Brokering System by using Selective encryption, Vigenere Cipher encryption and Selective Reverse Circle Cipher algorithm and Huffman Compression algorithm.

Future Work

1. System privacy can be enriched with waste token generation policy.
2. We can implement this system in Distributed Information brokering system
3. This technique use in law enforcement system.

6. Result

Figure 2 show that the graph of average request brokering time in ms of distributed and semantic web Information Brokering System Vs increase in keywords in which simulation result show semantic web approach requires less time.

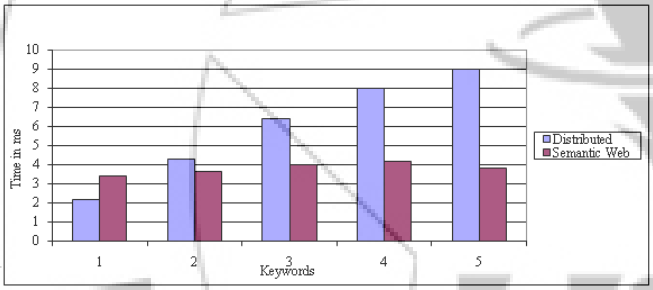

Figure 2: Average Request Brokering time

Acknowledgment

I express true sense of gratitude towards my project guide Prof. R. A. Mahajan, for his invaluable co-operation and guidance that she gave methroughout my project. I like to specially thank our Head of department Prof. G. M. Bhandari and P.G coordinator Prof. A. C. Lomte for inspiring me and providing me all the lab facilities, which made project work very convenient and easy. I would also like to express my appreciation and thanks to JSCOE principal Dr. D. M. Yadav and all my friends who knowingly or unknowingly have assisted me throughout my hard work.

References

[1] W. Bartschat, J. Burrington-Brown, S. Carey, J. Chen, S.Deming, and S. Durkin, "Surveying the RHIO landscape: A description of current {RHIO} models, with a focus on patient identification," *J. AHIMA*, vol. 77, pp. 64A–64D, Jan. 2006.

[2] A. P. Sheth and J. A. Larson, "Federated database systems for managing distributed, heterogeneous, and autonomous databases," *ACMComput. Surveys (CSUR)*, vol. 22, no. 3, pp. 183–236, 1990.

[3] Encryption Policies for Regulating Access to Outsourced Data, Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, Pierangela Samarati, *ACM Transactions on Database Systems, Vol. 35, No. 2, Article 12, Publication date: April 2010.*

[4] Dynamic and efficient key management for access hierarchies, Mikhail J. Atallah, Marina Blanton, Nelly Fazio, Keith B. Frikken, *ACM Transactions on Information and System Security, Vol. 12No. 3, Article 18, Pub. date: January 2009.*

[5] Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing, Shucheng Yu, Cong Wang, Kui Ren, Wenjing Lou, *INFOCOM, 2010 Proceedings IEEE , Publication Year: 2010.*

[6] Combining Attribute-Based and Access Systems, Malek, B. Miri, A., *Computational Science and Engineering, 2009. CSE '09. International Conference, Publication Year: 2009*

[7] Towards Efficient Over-Encryption in Outsourced Databases Using Secret Sharing, Shuai Liu ,Wei Li, Lingyu Wang, *New Technologies, Mobility and Security, 2008. NTMS '08.*

[8] Data protection in outsourcing scenarios: issues and directions, Pierangela Samarati, Sabrina De Capitani di Vimercati April 2010 ASIACCS '10**:** Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security

[9] [9] Ming Li , Shucheng Yu, Yao Zheng, Kui Ren*Member* and Wenjing Lou,"Scalable and Secure Sharing of Personal Health Records in Cloud Computing using Attribute-based Encryption," IEEE *Transactions* on Parallel And Distributed Systems,Jan. 2013.

[10] Virtual Private Data Website Available: http://oracle.com

[11] Ebenezer R.H.P. Isaac, Joseph H.R. Isaac and J. Visumathi, "Reverse Circle Cipher for Personal and Network Security", Information Communication and Embedded Systems (ICICES), 2013 International Conference on 21-22 Feb. 2013

[12] Vigenere cipher. Retrieved from http://en.wikipedia.org/wiki/Vigenère_cipher

[13] Fengjun Li, Bo Luo, Peng Liu, Dongwon Lee, and Chao-Hsien Chu, "Enforcing Secure and Privacy- Preserving Information Brokering in Distributed Information Sharing" , IEEE Transactions On Information Forensics And Security, Vol. 8, No. 6, 2013.

Author Profile

Ms Supriya Sankpal received her B.E. degree in Computer Engineering from SVPM College of engineering Malegoan (BK), Pune University, in 2008.Currently she is teaching as Senior lecturer with Department of Computer Engineering at JSPM's Jayawantrao Sawant Polytechnic, Pune, Maharashtra, India. since June 2009.In her post graduate course she is doing research work on Enforcing Privacy-Preserving Information Brokering in Semantic Web.

Prof. Rupali Mahajan M.E (CSE).She is working as an Asst. Professor in Information Technology Department of JSPM's Bhivarabai Sawant Institute of Technology & Research (for Women), Wagholi, Pune, Maharashtra, India