High-Dimensional Confidential Data Mash up using Service- Oriented Architecture

Pradeep Gurunathan, N. Ishwarya, V. Sridevi, C. Nandhini, S. Deepalakshmi

Abstract -- Mash up is integrating different service providers to expertise and to deliver highly customizable services to their customers. Simply joining multiple private data sets together would reveal the sensitive information to the other data providers. The integrated (mash up) data could potentially sharpen the identification of persons and therefore, expose their person-specific sensitive information that was not available before the mash up. The mash up data from multiple sources often contains many data attributes. When enforcing a established privacy model such as K-anonymity, the high-dimensional data would asssit from the problem known as the curse of high dimensionality, resulting in ineffective data for further data analysis. In this paper, we introduced a new algorithm called Modified privacy preserving high dimensional confidential (MPHDC) mash up algorithm to provide the high dimensional security to the user from the data provider.

Keywords: Confidential mash up, High dimensionality, Mash up service, etc.,

I. INTRODUCTION

MASHUP service is a web technology that combines various information from multiple sources into a single web application. An example of a successful mash up application is the combination of real estate information into Google Maps, which allows users to browse on the map for properties that satisfy their specified requirements. Developers create mashups by combining components of existing Web sites and applications. Mashups combine views, data, and logic from existing Web sites or applications to create novel applications that focus on situational and passing problems. In this paper, we focus on data mash up, a special type of mash up application that aims at integrating data from multiple data providers depending on the service request from a user (a data beneficiary). An information service request can be a common count statistic task or a stylish data mining task such as classification analysis. Conceptually, mashups are simply new Web applications that repurpose alive Web data and APIs.

Manuscript received on April, 2013.

Pradeep Gurunathan, Professor& Head, Department of Information Technology, A.V.C College of Engineering, Mannamapandal, Mayiladuthurai, Nagapattinam District, Tamilnadu, India.

Ms.N.Ishwarya, B.Tech. Student, Department of Information Technology, A.V.C College of Engineering, Mannamapandal, Mayiladuthurai, Nagapattinam District, Tamilnadu, India.

Ms.V.Sridevi, B.Tech. Student, Department of Information Technology, A.V.C College of Engineering, Mannamapandal, Mayiladuthurai, Nagapattinam District, Tamilnadu, India.

Ms.C.Nandhini, B.Tech. Student, Department of Information Technology, A.V.C College of Engineering, Mannamapandal, Mayiladuthurai, Nagapattinam District, Tamilnadu, India.

Ms.S.Deepa Lakshmi, B.Tech. Student, Department of Information Technology, A.V.C College of Engineering, Mannamapandal, Mayiladuthurai, Nagapattinam District, Tamilnadu, India.

Well-structured mashups therefore include all three aspects of an equivalently well designed Web application, data models, views, and interaction controllers. Also, mashups often intervene between mixed providers Web APIs.
The advertisements are posted to the user account by the admin or the mash up coordinator depending upon the category of the user. In this paper we use one more special attribute in the registration form. Generally the social network websites registration form consist of some basic details during signup like Name, Age, Gender, Username etc., In this paper we propose hiding option is also enable for protecting the sensitive information.

II. RELATED WORK

Jackson and Wang[1] presented a special mechanism namely secure communication mechanism it provides a cross domain network requests and responses by means of client-side communication. The main goal of this work is to protecting the mashup controller from hateful or malicious code through web services. In contrast, this paper aims to conserve the privacy and information service of the mashup data. Jhingran[2] presented the fundamental transformation that is enchanting place on the web around information composition through Mashups. Our hypothesis creates one type of mashup fabric. i.e., a new class of integration technologies will emerge to serve the integration and composition tasks, and we call it an enterprise information mashup fabric. The mashup fabric consists of a framework that augments. Using this framework we can achieve the combination from various sources using various services. They have demonstrated an information mashup fabric that builds up information primitives for situational applications. B.C.M. Fung, K. Wang, and P.S. Yu [3] presented the privacy-preserving data mashup problem studied in this paper allows data providers to share data, not only the data mining outcome. In many applications, data distribution gives greater suppleness than result sharing because the data recipients can perform their required analysis and data exploration**.** L. Sweeney[4] presented **K**-anonymity is a special case of LKC-privacy with L = |QID| and C = 100%, where |QID| is the number of QID attributes in the data table. Confidence bounding is also a special case of LKC privacy with L = |QID| and K = 1. (α,k) k-anonymity is a special case of LKC-privacy with L = |QID|, K = k, and C =α One instantiation of ‘-diversity is also a special case of LKC-privacy with L = |QID|, K = 1, and C = 1/l. Thus, the data provider can still achieve the traditional models. P. Samarati and L. Sweeney [5]presented the notation of K-anonymity, A data holder can often identify attributes in their data that also appear in outside sources, and these attributes are candidates for linking.

We term them, quasi-identifiers, and it is essentially the combinations of these quasi-identifiers that must be protected for providing security. Y. Lindell and B. Pinkas [6]presented the Privacy-preserving data mining, it containing the problem of running data mining algorithms on confidential data that is not supposed to be revealed even to the party running the algorithm. To achieve Secure Multiparty Computation the each cycle needs Background and Definitions for each process. The field of secure multiparty computation deals with the question of how to securely compute functionality, but does not ask the question of whether the functionality should be computed in the first place or not**.** Z. Yang, S. Zhong, and R.N. Wright [7] presented one technique called randomization technique. This is the solution for the previous problem.The problem is the scenario in the field of data miner survey, a large number of customers to learn classification rules on their data, while the sensitive attribute of these clients need to be protected. They presented a simple cryptographic approach that is efficient even in a many-customer setting, provides well-built privacy for each customer, and does not drop any accuracy as the price of privacy. The problem is solved using randomization techniques, and then there is a transaction between privacy and accuracy**.**

III.EXISTING SCENARIO

A data mash up application can help ordinary users explore new knowledge; it could also be misused by adversaries to reveal sensitive information that was not available before the mash up. High dimensionality is a critical obstacle for achieving effective data mash up because the integrated data from multiple parties usually contain many attributes. Enforcing traditional K-anonymity on high-dimensional data will result in significant information loss.

We first define the LKC-privacy model and the information service measure on a single data table, then extend it for privacy-preserving high-dimensional data mashup from multiple parties.

a. Isolation Measure
   i. Record linkage
   ii. Attribute linkage

b. Service Measure
   i. Service Measure for Classification Analysis.
   ii. Service Measure for General data analysis.

b. Privacy-Preserving High-Dimensional Data Mashup

Consider n data providers(y) and they are having own data table (T_y).T_y is the type of attribute and it also having 4 parameters namely { QID (Quasi-Identifying attribute), UID(User identification),S(Sensitive Information),Class } for the each data providers. These parameters are considered as the same set of records for different data providers. UID and Class are the shared attributes among all data providers. In our system for example we considered two data providers as y and z and Their QID and S factor is QID_y, QID_z,, S_y, S_y respectively. The privacy preserving algorithm helps us to check the values of these factors. If Quasi factors of y and z are different means the sensitive information cannot be accessed between them. This algorithm only provides the information about the user when only if the UID information is match with the another data provider's UID value. This algorithm also provides the minimal information with the help of LKC privacy requirement on mashup table. Using this minimal information the mashup coordinator chooses the type of details about the user.i.e whether the information is local or global. If its local means all the values of QID_j are known by the one provider, else it is declared as global.

IV. PROPOSED SYSTEM

The mash up coordinator receives an information service request from the data recipient and establishes connections with the data providers who can contribute their data to fulfill the request.

The mash up coordinator executes the privacy-preserving algorithm to integrate the private data from multiple data providers and to deliver the final mash up data to the data receiver. Note that our proposed solution does not require the mash up coordinator to be a trusted party.

Though the mash up coordinator manages the entire mash up service, our solution guarantees that the mash up coordinator does not gain more information than the final mash up data, thereby protecting the data privacy of every participant by using hide details option in the starting phase of the process. The mash up coordinator can be any one of the data providers or an independent party. This makes our architecture realistic for the reason that a trusted party is not always available in real-life mash up scenarios. In our proposed work we create two social networks. Using these networks we provide the high dimensional security at the registration phase itself and also provide the advertisement to the data recipients depending upon the user category.

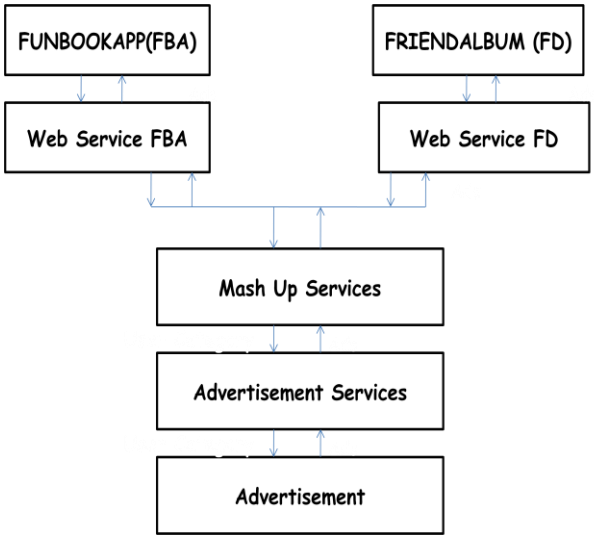

Figure 1: Work Flow Of The MPHDC Mashup Process

IV. ARCHITECTURAL APPROACH

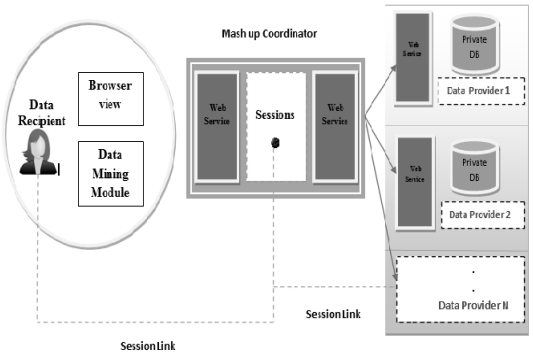

Figure 2: Service-oriented architecture for MPHDC Mashup

Figure 2 describes the architecture design of the privacy preserving confidential data mash up model. The data recipient containing the browser view model and the data mining model. Data mining model is used to extract the details of the customer for integration. The mash up coordinator containing the web services and session. Each data holder must connect with the session of the mash up coordinator part.web services provides the different kind of services related with the aim of the process and stored in the private database.

The data mash up process can be divided into two phases. In Phase I, the mash up coordinator receives the service request from the data recipient and establishes connections with the data provider who can contribute their data to fulfill the request. In Phase II, the mash up coordinator executes the privacy-preserving algorithm to integrate the private data from multiple data providers and to deliver the final mash up data to the data receiver. Note that our proposed solution does not require the mash up coordinator to be a trusted party. Though the mash up coordinator manages the entire mash up service, our solution guarantees that the mash up coordinator does not gain more details than the final mash up data, thereby shielding the data privacy of every participant. The mash up service have the following merits. They are,

1. Increase productivity
2. Increase innovation
3. Improve data security
4. Reduce burden to IT departments/increase freedom for business users from IT
5. Increase standardization across the enterprise
6. Bring an "App Store" model approach to development, vs. big-bang project.

A. Modified Phdc Mash-Up Algorithm

We are using MODIFIED PHDC MASH up algorithm to evaluate the impact on classification quality. A service-oriented architecture (SOA) that describes the communication paths of all participating party, followed by a privacy-preserving high-dimensional confidential data mash up algorithm that can efficiently identify a suboptimal resolution for the problem.SOA is an architectural model for developing and integrating heterogeneous information systems with strict message-driven communication model. Following the SOA design principles, the resulting system has several attractive properties including interoperability and loosely coupling. Interoperability means capability of allowing platform-independent design of the system components based on a common understanding of service component and interfaces. Loosely coupling refers to the capability of minimizing dependencies among the system components and therefore, improving the overall elasticity, scalability, and fault tolerance of a system. In this paper, we described data sources can be dynamically composed to serve new mashup requests depending on the data analysis tasks and privacy requirements. SOA having the capabilities of interoperability and loosely coupling has become a natural choice to tackle the heterogeneity of different potential data providers.

1:initialize T_g to embrace one testimony containing topmost values,
2: initialize $UCut_i$ to embrace only topmost values and update IsValid(s) for every s ε $UCut_i$
3: while v ε $UCut_i$ s.t. IsValid(s) do
4: find the local winner β that has the highest Score(β),
5: communicate Score(β) with provider B to determine the global winner z,
6: if the winner z is local then
7: specialize z on T_g,
8: instruct provider B to specialize z,
9: else
10: wait for the instruction from providerB.
11: specialize z on T_g using the instruction,
12: end if
13: replace w with child(z) in the local copy of $UCut_i$,
14: update Score(s) and IsValid(s) for every candidate s ε $UCut_i$,
15: end while
16: return T_g and $UCut_i$

Figure 3: Algorithm PHDC Mashup for Provider A (Same as Provider B)

The nature of the top-down approach implies that Tg is always more general than the final mash up table and therefore, does not violate necessities. At each iteration, the data provider cooperate to perform the same identified specialization by communicating some count statistics information that satisfies necessities. Below, we describe the key steps: find the winner contender (Lines 4-5), perform the winner specialization (Lines 7-11), and update the score and status of contenders (Line 14). For provider A, a local attribute refers to an attribute from T_A.

V. RESULTS AND DISCUSSION

In this information structural design we create the two social networks. The web services are created by the help of the user category that are enrolled by the user during registration. In registration they give their entire information to the social network and then these details are submitted to the data providers .In existing information the entire details about the user is viewed by the all data providers. So it may be threaten to the user to overcome this problem here we use the mash up algorithm with k-anonymity model for providing security.

We apply a data mashup function for the online advertising industry in social networks, and generalize their privacy and information requirements to the problem of privacy-preserving data mashup for the purpose of joint data analysis on the high-dimensional data.

VI. CONCLUSION

Our future work lies in deploying Mashups in a real forum to better evaluate its performance. Further, we will consider the factors that affect the quality of service such as the factors of the mashup services and securing the data depending upon the user category. These are the side effects of our project. We will investigate how to deal with the side effects in our future work.

REFERENCES

1. C. Jackson and H.J. Wang (2007), "Subspace: Secure Cross-Domain Communication for Web Mashup," Proc. 16th Int'l Conf. World Wide Web, pp. 611-620.
2. Jhingran, "Enterprise Information Mashups: Integrating Information, Simply," Proc. 32nd Int'l Conf. Very Large Data Bases, pp. 3-4, 2006.
3. B.C.M. Fung, K. Wang, and P.S. Yu (May 2007), "Anonymizing Classification Data for Privacy Preservation," IEEE Trans. Knowledge and Data
4. L. Sweeney (2002) "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression," Int'l J. Uncertainty, Fuzziness, and Knowledge-Based Systems.
5. P. Samarati and L. Sweeney(June 1998), "Generalizing Data to Provide Anonymity when Disclosing Information," Proc. 17th ACM SIGACT-SIGMOD-SIGART Symp. Principles of Database Systems, p. 188
6. Y. Lindell and B. Pinkas(2009), "Secure Multiparty Computation for Privacy-Preserving Data Mining," J. Privacy and Confidentiality, vol. 1, no. 1, pp. 59-98
7. Z. Yang, S. Zhong, and R.N. Wright (2005), "Privacy-Preserving Classification of Customer Data without Loss of Accuracy," Proc. Fifth SIAM Int'l Conf. Data Mining, pp. 92-102

AUTHORS PROFILE

Pradeep Gurunathan, received his Master Degree in M.S University, Thirunelveli and pursuing his Doctoral Degree in Anna University, Chennai .Currently, he is working as Professor and Heading the Department of Information Technology in A.V.C College of Engineering, Mannamapandal, Mayiladuthurai, Nagapattinam District, Tamilnadu, India. He received various academic excellence awards and fetched various grants such as FDP, Seminar and MODROBS from AICTE, New Delhi. His area of interest includes Web Services, SOA, Internet Technologies and Cloud Computing.

Ms. N.Ishwarya, Currently doing her B.Tech Information Technology in A.V.C College of Engineering. Her area of interest embraces Web services, Web Desinging ,Multimedia and Network security.She presented her paper in conference and she participated in different technical activities.

Ms.V.Sridevi received her Diploma in Computer Engineering from A.V.C Polytechnic Mannampandal. Currently, she is doing B.Tech Information Technology in A.V.C College of Engineering. Her area of interest embraces Networking and pervasive computing.

Ms.C.NANDHINI received her Diploma in Computer Enggineering from A.V.C Polytechnic Mannampandal. Currently doing her B.Tech Information Technology in A.V.C College of Engineering. Her area of interest embraces Networking and Web Desiging.

Ms. S.DEEPA LAKSHMI received her Diploma in Computer Enggineering from A.V.C Polytechnic Mannampandal. Currently doing her B.Tech Information Technology in A.V.C College of Engineering. Her area of interest embraces Networking and debugging.