International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 3 Issue 11, November 2014

Implementation of SOA and mashup technology to maintain confidentiality of private data

Ms.Deepali V.Shinkar, Prof. S. Pratap Singh

*Abstract— — **Data Mashup is the process of integrating information from different service providers and putting together for various purposes. Privacy protection on private data in the following scenario: Multiple parties, each having a private data set, want a group of people organized for a joint purpose rule mining without disclosing their private data to other parties. Because of the interactive nature among parties, developing a secure framework to achieve such a computation is both challenging and desirable. This system integrates various networking sites with common SOA framework to give the same type of services from a single data provider. The integrated data could potentially sharpen the identification of persons and therefore expose their person specific sensitive information that was not available before the mashup.In this paper we study how to integrate and secure sensitive data which minimizes the privacy threat with the help of data mashup and propose a service-oriented architecture for privacy-preserving data mashup.The mashup data from multiple sources often contains many data attributes. We use technique such as a new privacy model called LKC-privacy to overcome the challenges and present centralized anonymization algorithms to achieve LKC-privacy for multiple data providers. Experiments demonstrate that our centralized anonymization algorithms can effectively retain the essential information in anonymous data for data analysis and is scalable for anonymizing large datasets. Our proposed method is effective for simultaneously preserving both privacy and information usefulness**.*

*Index Terms—**Privacy Protection, centralized anonymization, data mashup, service oriented architecture ,curse of high-dimesionaity***

INTRODUCTION

A **mashup**, in web development, is a web page, or web application, that uses content from more than one source to create a single new service displayed in a single graphical interface. For example, you could combine the addresses and photographs of your library branches with a Google map to create a map mashup. The term implies easy, fast integration, frequently using open application programming interfaces (API) and data sources to produce enriched results that were not necessarily the original reason for producing the raw source data.

The main characteristics of a mashup are combination, visualization, and aggregation. It is important to make existing data more useful, for personal and professional use. To be able to permanently access the data of other services, mashups are generally client applications or hosted online.

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. An information service request can be a common count statistic task or a stylish data mining task such as classification analysis. Mashup often interface between mixed providers Web APIs.However, there is a potential privacy risk because of the possibility of having sensitive information revealed which was impossible or not obvious before the integration.

We generalize their problem described as follows. A loan company A,a bank B,a customer C observe different sets of attributes about the same set of individuals identified by the common key unique identifier number (UID), e.g., TA(Sensitive value,Gender,Work-class,Hours-per-week), TB(UID,Job,Age,Race),TC(UID,Education,Salary).These data providers want to implement a data mashup application that integrates their data to support better decision making such as loan or credit limit approval, which is basically a data mining task on classification analysis.

In addition to companies A, B, C their partnered credit card company D also has access to the data mashup application, so all three companies A, B, C,D are data recipients of the final integrated data. Companies A,B,C have two privacy concerns. First, simply joining TA,TB,TC would reveal the sensitive information to the other party. Second, even if TA,TB,TC individually do not contain person-specific or sensitive information, the integrated data can increase the possibility of identifying the record of an individual.

SHARED		PROVIDER A		PROVIDER B		PROVIDER C	
Uid	Class	Sensitive	Gender	Job	Age	Education	City
1	Y	s1	M	Lawyer	39	Bachelors	Mumbai
2	N	s1	M	Lawyer	50	Bachelors	Kolkatta
3	Y	s2	M	Lawyer	38	Doctorate	Shimala
4	N	s2	M	Janitor	53	11th	Pune
5	N	s1	F	Lawyer	28	Bachelors	Chennai
6	Y	s2	F	Doctor	37	Masters	Banglor
7	N	s2	F	Carpenter	49	9th	Ludiyanna
8	N	s2	M	Doctor	52	Masters	Dehli
9	N	s2	F	Janitor	31	10th	Panaji
10	Y	s2	M	Lawyer	42	Bachelors	Patana
11	Y	s1	M	Technician	37	12th	Mumbai

Table I : INTEGRATED RAW DATA TABLE

Manuscript received Nov.2014
Ms,Deepali Shinkar Department of Computer & Student of IOK,COE,Pune.
Prof.S.Pratap Singh Department of Computer &Assist. Professor of IOK,COE,Pune.

ISSN: 2278 – 1323

From Table 1 After integrating the three tables (by matching the UID field), the male,lawyer,doctorate,shimala on (Sex, Job,education,city) becomes unique, therefore, vulnerable to be linked to sensitive information such as Salary. To prevent such linking, we can generalize T and Lawyer,technician,Carpenter to Professional so that this individual becomes one of many female or male professionals. No information is lost as far as classification is concerned because Class does not depend on the distinction of Technician,Carpenter and Lawyer.

.

I. RELATED WORK

Motivated by the privacy concerns on data mining tools, a research area called privacy-preserving data mining (PPDM) emerged in 2000 [1,2]. The initial idea of PPDM was to extend traditional data mining techniques to work with the data modified to mask sensitive information. The key issues were how to modify the data and how to recover the data mining result from the modified data. The solutions were often tightly coupled with the data mining algorithms under consideration.

A number of techniques have been proposed for modifying or transforming the data in such a way so as to preserve privacy. A privacy threat occurs when an adversary is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call these record linkage, attribute linkage, and table linkage, respectively.

In Randomized method noise is added to the data in order to mask the attribute values of records [1,2].Therefore, techniques such as Additive perturbation, matrix perturbation, data swapping are designed to derive aggregate distributions from the perturbed records. The k-anonymity techniques is record linkage model [4], we reduce the granularity of representation of these pseudo-identifiers with the use of techniques such as generalization and suppression. Datafly system [8] and μ-Argus system [9] use generalization to achieve K-anonymity. Mohammed et al. [10] propose a top-down specialization algorithm to securely integrate two vertically partitioned distributed data tables to a K-anonymous table, and further consider the participation of malicious parties in [11].l-Diversity technique is record linkage and attribute linkage model.It provides privacy even when the data publisher does not know what kind of knowledge is possessed by the adversary. Concept of intra-group diversity of sensitive values is promoted within the anonymization scheme [6].

Typically, such methods reduce the granularity of representation in order to reduce the privacy. This reduction in granularity results in someloss of effectiveness of data management or mining algorithms. This is the natural trade-off between information loss and privacy. Jiang and Clifton [14][15] propose a cryptographic approach.Yang et al. [16] develop a cryptographic approach to learn classification rules from a large number of data providers while sensitive attributes are protected. The problem can be viewed as a horizontally partitioned data table in which each transaction is owned by a different data provider. The output of their method is a classifier, but the output of our method is an anonymous mashup data that supports general data analysis or classification analysis[17].

Secure multiparty computation (SMC) [23], [24] on the other hand, allows sharing of the computed result (e.g., a classifier), but completely prohibits sharing of data. Output perturbation techniques discuss privacy with respect to the information released as a result of querying a statistical database by some external entity.

Mohammed et al. [26] extend the work to address the problem of high-dimensional anonymization of or the healthcare sector using LKC-privacy[4]. All these works consider a single data source; therefore, data mashup is not an issue. Recently, Mohammed et al. [27] propose an algorithm to address the horizontal integration problem,while our paper addresses the vertical integration problem.

Trojer et al. [28] present a service-oriented architecture for achieving K-anonymity in the privacy preserving data mashup scenario. Our paper is different from these previous works [12], [13], [10], [11], [28] in two aspects. First, our LKC-privacy model provides a stronger privacy guarantee than K-anonymity because K-anonymity does not address the privacy attacks caused by attribute linkages, as discussed in survey table Second, our method can better preserve information utility in high-dimensional mashup data. High dimensionality is a critical obstacle for achieving effective data mashup because the integrated data from multiple parties usually contain many attribute. Our privacy model resolves the problem of high dimensionality.

II. PROBLEM STATEMENT

We study the privacy threats caused by data.The integrated table must satisfy both the following anonymity and information requirements:

- **Anonymity Requirement**:

The integrated table has to satisfy k-anonymity A data table T satisfies k-anonymity if every combination of values on QID is shared by at least k records in T , where the quasi-identifier (QID) is a set of attributes in T that could potentially identify an individual in T , and k is a user-specified threshold. k-anonymity can be satisfied by generalizing domain values into higher level concepts. In addition, at any time in the procedure of generalization, no party should learn more detailed information about the other party other than those in the final integrated table.

- **Information Requirement:**

The generalized data should be as useful as possible to classification analysis. Generally speaking, the privacy goal requires masking sensitive information that is specific enough to identify individuals, whereas the classification goal requires extracting trends and patterns that are general enough to predict new cases. If generalization is carefully performed, it is possible to mask identifying information while preserving patterns useful for classification.

In addition to the privacy and information requirements, the data mashup application is an online web application. The user dynamically specifies their requirement and the system is expected to be efficient and scalable to handle high volumes of data.

ISSN: 2278 – 1323

Privacy-preserving data mashup Given multiple private tables T1, . . . , Tn, a joint anonymity requirement {“QID1, k1", . . . , QIDp, kp"}, and To generalize T, a taxonomy tree is specified for each categorical attribute in UQIDj. For a numerical attribute in UQIDj , a taxonomy tree can be grown at runtime, where each node represents an interval, and each non-leaf node has two child nodes representing some optimal binary split of the parent interval. The algorithm generalizes a table T by a sequence of specializations starting from the top most general state in which each attribute has the top most value of its taxonomy tree. A specialization, written v →child(v), where child(v) denotes the set of child values of v, replaces the parent value v with the child value that generalizes the domain value in a record.A taxonomy tree for each categorical attribute in QIDj , the problem of privacy-preserving data mashup is to efficiently produce a generalized integrated table T such that

1. T satisfies the joint anonymity requirement,
2. Contains as much information as possible for classification, and
3. Each party learns nothing about the other party more specific than what is in the final generalized

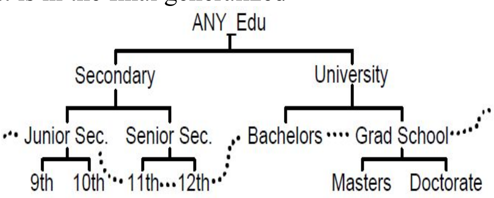


Fig. 1.Taxonomy Tree and QIDs.

Sensitive	Gender	Job	Age	Education	City
s1	M	Professional	(30-60)	Bachelors	West
s1	M	Professional	(30-60)	Bachelors	East
s2	M	Professional	(30-60)	Grand school	North
s2	M	Non-Technical	(30-60)	Senior-sec	West
s1	F	Professional	(10-30)	Bachelors	South
s2	F	Professional	(30-60)	Grand school	South
s2	F	Technical	(30-60)	Junior-Sec	North
s2	M	Professional	(30-60)	Grand school	North
s2	F	Non-Technical	(10-30)	Junior-Sec	West
s2	M	Professional	(30-60)	Bachelors	East
s1	M	Technical	(30-60)	Senior-sec	West

TABLE II ANONYMOUS MASHUP DATA(L=2,K=2,C=50%)

In case all QIDs are locals, we can generalize each table TA,TB,TC independently, and join the generalized tables to produce the integrated data. However, if there are global QIDs, global QIDs are ignored in this approach. Further generalizing the integrated table using global QIDs does not work because the requirement (3) is violated by the intermediate table that contains more specific information than the final table. It may seem that local QIDs can be generalized beforehand. However, if a local QIDi shares some attributes with a global QIDg, the local generalization ignores the chance of getting a better result by generalizing QIDg first, which leads to a sub-optimal solution. A better strategy is generalizing shared attributes in the presence of both QIDi and QIDg. Similarly, the generalization of shared attributes will affect the generalization of other attributes in QIDi, thus, affect other local QIDs that share an attribute with QIDi. As a result, all local QIDs reachable by a path of shared attributes from a global QID should be considered in the presence of the global QID.

III. MATHEMATICAL MODEL

Consider n data providers Provider 1 ……….Provider n where each provider y owns a private table T (UID,QIDy,Sy,Class) over same set of records.UID and Class are shared attributes among all data providers. QIDy is a set of quasi-identifying attributes and Sy is set of sensitive values owned by provider y.

QIDy ∩ QIDz and Sy ∩ Sz for any 1<=y, 1<=z.

These providers agree to release “minimal information” to form a mashup table T (by matching the UID) for conducting general data analysis or a joint classification analysis. The notion of minimal information is specified by an LKC-privacy requirement on the mashup table. A QIDj is local if all attributes in QIDj are owned by one provider; otherwise, it is global.

NP is the class of problems which have efficient verifiers i.e. there is a polynomial time algorithm that can verify if a given answer is correct.

The algorithm generalizes a table T by a sequence of specializations starting from the top most general state in which each attribute has the top most value of its taxonomy tree. A specialization, written v →child(v), where child(v) denotes the set of child values of v, replaces the parent value v with the child value that generalizes the domain value in a record.

- A specialization is valid if the specialization results in a table satisfying the anonymity requirement after the specialization.
- A specialization is beneficial if more than one class are involved in the records containing.

The verifier V gets two inputs,

- T: the generalized table input
- LKC is suggested input

One method is computing Score, which measures the goodness of a specialization with respect to privacy preservation and information preservation.
The effect of a specialization v → child(v) can be summarized by information gain, denoted InfoGain(v), and anonymity loss, denoted AnonyLoss(v), due to the specialization. Our selection criterion is to favor the specialization v that has the maximum information gain per unit of anonymity loss.

$$Score(V) = \frac{\text{InfoGain(v)}}{\text{AnonyLoss(v)+1}} \tag{1}$$

We add 1 to AnonyLoss(v) to avoid division by zero.
InfoGain(v): Let T[x] denote the set of records in T generalized to the value x. Let freq(T[x]; cls) denote the number of records in T[x] having P the class cls.
Note that

$$|T[C]| = \Sigma_C \, |[c]|$$

Where c ε child (v).We have $\Sigma_c \frac{|T[c]|}{|T[V]|}$

$$InfoGain(v) = I(T(v) - \Sigma_c \frac{|T[c]|}{|T[V]|} \; I(T[c])$$

(2)

Where I(T[x]) is the entropy of T[x] :

ISSN: 2278 – 1323

$$I(T[x]) = \sum_{cls} \frac{freq(T[x],cls)}{|T[x]|} \; X \; \log_2 \frac{freq(T[x],cls)}{|T[x]|} \quad (3)$$

Intuitively, I(T[x]) measures the mix of classes for the records in T[x], and InfoGain(v) is the reduction of the mix by specializing v.

AnonyLoss(v): This is the average loss of anonymity by specializing v over all QIDj that contain the attribute of v:

$$AnonyLoss(v) = avg\{A(QID_i - QID_j)\} \quad (4)$$

where A(QIDj) and Av(QIDj) represents the anonymity before and after specializing v. Note that AnonyLoss(v) not just depends on the attribute of v; it depends on all QIDj that contain the attribute of v. Hence, avg{A(QIDj) Av(QIDj)} is the average loss of all QIDj that contain the attribute of v.

IV. PROPOSED ARCHITECTURE AND PROTOCOL

We present a service-oriented architecture (SOA) that describes the communication paths of all participating parties, followed by a privacy-preserving protocol that can efficiently identify a suboptimal solution for the above described problem.

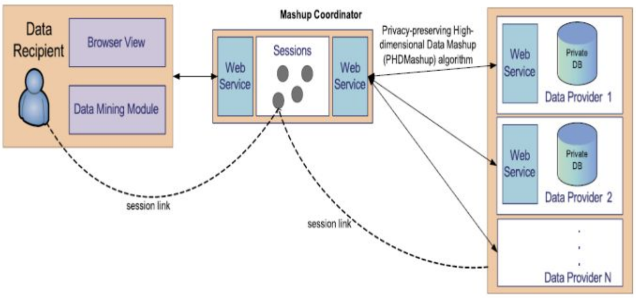


Figure2.Service-oriented architecture for privacy-preserving data mashup

Mentioning to the architecture shown in Fig. 2, the data mashup process can be divided into two phases.

- **Phase I: Session Establishment**

The mashup 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 objective of Phase I is to establish a common session context between the data recipient and the contributing data providers. An operational context is successfully established by proceeding through the steps of data recipient authentication, contributing data providers identification, session context initialization, and common requirements negotiation.

- **Phase II: Privacy-Preserving Protocol**

After a common session has been established among the data providers, the mashup coordinator initiates the privacy preserving data mashup protocol (PPMashup) and stays back. Upon the completion of the protocol, the mashup coordinator will receive an integrated table that satisfies both, the information and anonymity requirements. There are two advantages that the mashup coordinator does not have to participate in the PPMashup protocol. First, the architecture does not require the mashup coordinator to be a trusted entity. The mashup coordinator only has access to the final integrated k-anonymous data. Second, this setup removes the computation burden from the mashup coordinator, and frees up the coordinator to handle other requests.

One major contribution of this paper is to extend a single party anonymization algorithm, called top-down specialization (TDS) [5], to a multiparty privacy-preserving data mashup to solve the problem of curse of high dimesionality.

Algorithm :Centralized algorithm for multiple data providers n executed by mashup co-ordinator

//Mashup Co-ordinator generates a new session id for synchronizing n provider instances of one session & sends to all n providers.

1. Initialize UCuti to include only topmost values and update isvalid(v) for every v ε UCuti //Every provider initialize Tg to include one record containing topmost values
2. while some candidate v ε UCuti s.t.Isvalid(v) do
3. Find Local winner(α) that has highest score(α) co-ordinator gathers local winners of all providers & then calculate global winner w.
4. if the winner w is local then instruct the local winner provider to do specialization on winner value of UCuti
5. else
6. Wait for the instruction from local winner of provider x specialization w on Tg
7. end if
8. Replace w with child(w) in local copy of UCuti
9. Update score(v) and Isvalid(v) for every candidate v v ε UCuti //This process repeat until all co-ordinators doesn't have any valid local winner
10. end while //Then co-ordinator instructs to resume finding local winner procedure to all providers
11. Display Final value as Tg and UCuti,// After this co-ordinator collects data from all providers in UCuti format

Centralized anonymization algorithm for multiple parties At each iteration, the data providers cooperate to perform the same identified specialization by communicating some count statistics information that satisfies requirement section 3.We describe the key steps: find the winner candidate (Lines 4-5), In Line 9, each party should communicate with all the other parties for determining the winner. Perform the winner specialization (Lines 7-9), Similarly, in Line 9, the party holding the winner candidate should instruct all the other parties and in Line 6, a party should wait for instruction from the winner party.and update the score and status of candidates (Line 9).

ISSN: 2278 – 1323

TABLE3:ADULT DATA SET

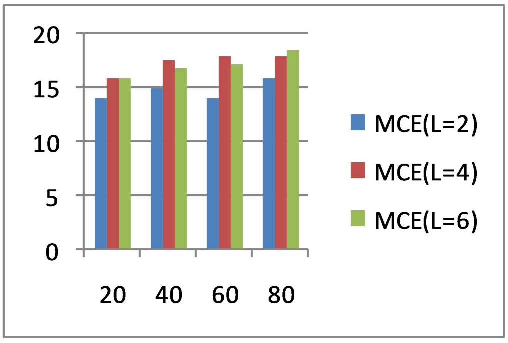
Attribute	Type	Numerical-range	
		# Levels	# Leaves
Age	Numerical	17-90	
Education	Categorical	16	5
Race	Categorical	2	2
Sex	Categorical	2	2
Martial-status	Categorical	7	4
Native city	Categorical	20	5
Hours-per-week	Numerical	13-99	
Work-class	Categorical	8	5
Occupation	Categorical	14	3

V. RESULT

We implement the proposed PHDMashup in a distributed web service environment. Each data provider is running on an Intel Core2 Quad Q6600 2.4 GHz PC with 2 GB RAM connected to a LAN. To evaluate the benefit of data mashup for joint data analysis,Due to the privacy agreement, we cannot use the raw data from the social network companies for experiments, so we employ the de facto benchmark census data set Adult, which is also a real-life data set, to illustrate the performance of our proposed architecture and algorithm. The Adult data set has six numerical attributes, eight categorical attributes, and a binary Class attribute representing two income levels _50 K or >50 K. Table 3 describes each attribute. It contains 45,222 records after removing records with missing values. We model a 3-data provider scenario with three private tables TA ,TB, TC as follows: TA contains the first 4 attributes, and TB contains 5 attributes and Tc contains remaining 5 attributes. A common UID is added to three tables for joining. The taxonomy trees for numerical and categorical attributes are presents.

- **Benefits of Mashup**

Lower classification error means better data quality. We collect two types of classification errors from the testing set: Mashup Classification Error (MCE) is the error on the mashup data produced by our Centralized Anonymization algorithm. for multiple data providers.Source error (SE) is the error on individual raw data table without generalization. SE forTA, denoted by SE for TA , is 19 percent and SE for TB, denoted by SE for TC, is 18 percent. SE _MCE measures the benefit of data mashup over individual private table.



Threshold K

Fig. 3.Benefits of mashup(C=20%)

Fig. 3depicts the MCE for the adversary's prior knowledge L = 2, L = 4, and L = 6 with confidence threshold C =20% and anonymity threshold K ranging from 20 to 100. The benefit decreases as L increases because more generalization is required in order to thwart the linkage attacks. In practice, the benefit is more than the accuracy consideration because our method allows the participating data providers to share data for joint data analysis, rather than sharing a classifier from each provider.

VI. CONCLUSIONS

In this paper we studied how to integrate and secure sensitive data which minimizes the privacy threat with the help of data mashup and propose a service-oriented architecture for privacy-preserving data mashup whereas the integrated data still retains the essential information for supporting general data exploration or a specific data mining task, such as classification analysis.

REFERENCES

[1]R. Agrawal and R. Srikant. Privacy preserving data mining. In Proc. of ACM International Conference on Management of Data (SIGMOD), pages 439–450, Dallas, Texas, May 2000.

[2] Agrawal D. Aggarwal C. C. On the Design and Quantification of Privacy-Preserving Data Mining Algorithms. *ACM PODS Conference*, 2002.

[3] Aggarwal C. C.: On Randomization, Public Information and the Curse of Dimensionality. *ICDE Conference*, 2007.

[4] Samarati P.: Protecting Respondents' Identities in Microdata Release. IEEE Trans. Knowl. Data Eng. 13(6): 1010-1027 (2001).

[5] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati k-anonymity Springer US, Advances in Information Security (2007)

[6] Machanavajjhala A, Gehrke J, Kifer D (2006). `*l*-diversity: Privacy beyond *k*-anonymity. In Proc. of the International Conference on Data Engineering ICDE'06), Atlanta, GA, USA.

[7]C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu. Tools for privacy preserving distributed data mining. ACM SIGKDD Explorations Newsletter, 4(2):28–34, December 2002.

ISSN: 2278 – 1323

[8] L. Sweeney, "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression," Int'l J. Uncertainty, Fuzziness, and Knowledge-Based Systems, vol. 10, no. 5, pp. 571-588, 2002.

[9] A. Hundepool and L. Willenborg, "μ- and τ-Argus: Software for Statistical Disclosure Control," Proc. Third Int'l Seminar Statistical Confidentiality, 1996.

[10]N. Mohammed, B.C.M. Fung, K. Wang, and P.C.K. Hung, "Privacy-Preserving Data Mashup," Proc. 12th Int'l Conf. Extending Database Technology (EDBT), pp. 228-239, Mar. 2009.

[11] N. Mohammed, B.C.M. Fung, and M. Debbabi, "Anonymity Meets Game Theory: Secure Data Integration with Malicious Participants," Int'l J. Very Large Data Bases, vol. 20, pp. 567-588, 2011

[12] W. Jiang and C. Clifton, "Privacy-Preserving Distributed k- Anonymity," Proc. 19th Ann. IFIP WG 11.3 Working Conf. Data and Applications Security, pp. 166-177, Aug. 2005.

[13] W. Jiang and C. Clifton, "A Secure Distributed Framework for Achieving k-Anonymity," J. Very Large Data Bases, vol. 15, no. 4, pp. 316-333, Nov. 2006.

[14] B.C.M. Fung, K. Wang, and P.S. Yu, "Anonymizing Classification Data for Privacy Preservation," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 5, pp. 711-725, May 2007.

[15] W. Jiang and C. Clifton, "A Secure Distributed Framework for Achieving k-Anonymity," J. Very Large Data Bases, vol. 15, no. 4, pp. 316-333, Nov. 2006.

[16 Z. Yang, S. Zhong, and R.N. Wright, "Privacy-Preserving Classification of Customer Data without Loss of Accuracy," Proc. Fifth SIAM Int'l Conf. Data Mining, pp. 92-102, 2005.

[17] Benjamin C.M. Fung, Patrick C.K. Hung, Khalil Al-Hussaeni, "Service-Oriented Architecture for High-Dimensional Private Data Mashup IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 5, NO. 3, JULY-SEPTEMBER 2012

[18]P. Jurczyk and L. Xiong, "Privacy-Preserving Data Publishing for Horizontally Partitioned Databases," Proc. 17th ACM Conf. Information and Knowledge Management, Oct. 2008.

[19] P. Jurczyk and L. Xiong, "Distributed Anonymization: Achieving Privacy for Both Data Subjects and Data Providers," Proc. 23rd Ann. IFIP WG 11.3 Working Conf. Data and Applications Security (DBSec)2009
(DBSec), 2009.

[20] A. Jhingran, "Enterprise Information Mashups: Integrating Information, Simply," Proc. 32nd Int'l Conf. Very Large Data Bases, pp. 3-4, 2006.

[21]G. Wiederhold, "Intelligent Integration of Information," Proc. ACM Int'l Conf. Management of Data (SIGMOD), pp. 434-437, 1993.

[22] R. Agrawal, A. Evfimievski, and R. Srikant, "Information Sharing Across Private Databases," Proc. ACM Int'l Conf. Management of Data (SIGMOD), 2003.

[23] O. Goldreich, Foundations of Cryptography: Vol. II Basic Applications. Cambridge Univ. Press, 2004.

[24] Y. Lindell and B. Pinkas, "Secure Multiparty Computation for Privacy-Preserving Data Mining," J. Privacy and Confidentiality, vol. 1, no. 1, pp. 59-98, 2009

[25] C. Jackson and H.J. Wang, "Subspace: Secure Cross-Domain Communication for Web Mashups," Proc. 16th Int'l Conf. World Wide Web, pp. 611-620, 2007.

[26] N. Mohammed, B.C.M. Fung, P.C.K. Hung, and C. Lee, "Anonymizing Healthcare Data: A Case Study on the Blood Transfusion Service," Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 1285-1294, June 2009..

[27] N. Mohammed, B.C.M. Fung, P.C.K. Hung, and C.K. Lee, "Centralized and Distributed Anonymization for High-Dimensional Healthcare Data," ACM Trans. Knowledge Discovery from Data, vol. 4, no. 4, pp. 18:1-18:33, Oct. 2010.

[28] T. Trojer, B.C.M. Fung, and P.C.K. Hung, "Service-Oriented Architecture for Privacy-Preserving Data Mashup," Proc. IEEE Seventh Int'l Conf. Web Services, pp. 767-774, July 2009.

[29]C.C. Aggarwal, "On k-Anonymity and the Curse of Dimensionality," Proc. 31st Very Large Data Bases, pp. 901-909, 2005.

[30] B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM
Computing Surveys, vol. 42, no. 4, pp. 14:1-14:53, June 2010

[31]PRIVACY-PRESERVING DATA MINING : MODELS AND ALGORITHMS Edited by
CHARU C. AGGARWAL PHILIP S. YU **Kluwer Academic Publishers** London

[32] Introduction tonPrivacy-Preserving Data Publishing Concepts and Techniques Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Philip S. Yu A Chapman & Hall Book

ISSN: 2278 – 1323