Utilization of SOA and mashup technology to conserve secrecy of competitive data

Mr. Dinesh Patil [1], Prof Umesh Lilhore [2]

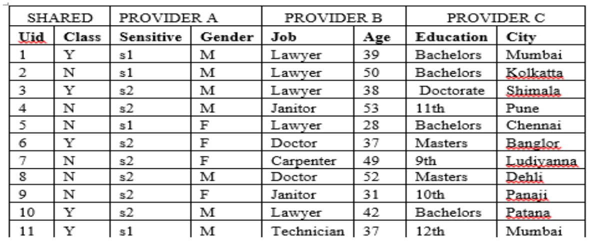
M. TechScholar Engineering, Department Computer Science [1,] Associate Prof Dept. of Computer Science Engineering [2,] , NRI-IIST Bhopal, M.P., State, India [1,] ,NRI-IIST Bhopal, M.P, State, India [2]
Email: dineshpatil83@gmail.com, umeshlilhore@gmail.com [2]

Abstract — Data Mashup is the Mechanizm of collecting Data from distinguishable service providers and swing together for various intentions. Privacy protection on private data in the next scenario: Multiple cooperators, each having a private data set, want a group of people organized for a combine. Target rule mining without disclosing their private data to other parties so the interactive nature in parties, Create and developing a secure framework to get such a computation is both make challenging and desirable. This system Combine 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 disclose their person specific sensitive information that was unavailable before the mashup.In this paper we study how to Combine and Protect sensitive data which loss the privacy threat with the help of data mashup and propose a service-oriented architecture to privacy-preserving in data mashup.The mash up data from Various sources often contains large data attributes. We use technique such as a latest privacy model is known as LKC-privacy to again come the remit and present centralized anonymization algorithms to get LKC-privacy for distict and many data providers. Experiments demonstrate that our centralized anonymization algorithms can effectively retain the essential information in inverse data for data analysis and is Countable for inversing large datasets. Our proposed method is useful for simultaneously preserving both privacy and information Active.

Keyword—Privacy Protection, centralized anonymization, data mashup, service oriented architecture, curse of high-dimensionality

1. INTRODUCTION

A mashup, is used web Technology, it contains web pages or web applications that apply content from multiple source to create a newer one service reflecting in a single one graphical interface. For example, you could combine the detail of residence and and photographs of your library Location with a Google map to create a map mashup. The concept implies simple, easy speedy collection, frequently taking open (API) application programming interfaces and data sources which generate better output that were not essential the main reason for generating the raw data source data. The real advantages characteristics of a mashup are collection, visualization, and accumulation. It is crucial to make previous data more applicable, for personal and professional work. To be enable to permanently and consistently availability of the data of another services, mashups are normally client applications software or hosted by online. Data mash up, a peculiar type of mash up application that aims at combining data from multiple data providers parasitic on the service request from a user. An information service request can be a common count statistic activity or a smart data mining task such as classification analysis. Mashup general convergence in between complex providers Web APIs.However, there is a prospective privacy risk because of the possibility of having critical information disclosed which was not possible or not obvious before the combine. We generalize their problem given as below. A loan Enterprise M,a bank N,a consumer K noticed various sets of attributes about the same set of individuals identified by the common key unique identifier number (UID), as like 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 collect and combine their own data to support exact decision making as like approval limit of loan . Which is basically a data mining task on classification analysis? In addition to companies M, N, K their partnered credit card company L also has authority the data mashup application, so all three Party M, N, K, L are data acquirer of the final collected data. Parties M, N, K have two privacy interests. First, simply joining TM, TN, TL would reveal the critical Data to the other party. Second, even if TM, TN, TL separately 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

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, so, compromising to be linked to critical Data such as Salary. To secure such linking, we can Expand T and Lawyer, technician, Carpenter to Professional so that this individual becomes one of multiple female or male participants. No Data is lost as far as classification is interest because Class does not depend on the distinction of Technician, Carpenter and Lawyer.

2. RELATED WORK

Intended by the privacy purpose on data mining tools, a research area familiar as privacy-preserving data mining (PPDM) come out in 2000 [1,2]. The Earlier concept of PPDM was to expand conventional data mining Method to work with the data modified to hide sensible information. The main problem was how to modify the data and how to recover the data mining result from the modified data. This solution was generally fixed up with the data mining algorithms which are under consideration. A multiple Procedure have been proposed for enlightening or fluctuation the data in such a way that to preserve privacy. A privacy threat happens when an opponent is able to link a record owner to a record in a publicized data table, to a sensible attribute in a publicized data table, or to the publicized data table itself. We call one by one these record , attribute and table linkage, respectively In Randomized method noise is adds to the data in order to that mask the attribute values of these records [1,2].Therefore, techniques such as Additive perturbation, matrix perturbation, data swapping have designed to give aggregate and average saturation from the perturbed records. The k-anonymity techniques is record linkage model [4], we minimize the granularity of each of representation of these pseudo-identifiers with the apply of techniques such as generalization and suppression. Data 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 combines two vertically partitioned distributed and saturates data tables for 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 given privacy even when the data publisher unknown about what kind of knowledge is presented by the adversary. The term of diversity of intra-group to be as sensitive values is promoted within the anonymization scheme [6].Mostly such methods minimize the granularity of demonstration in order to minimize the privacy. This deduction in granularity results in few loss of effectiveness and impression of data management and maintenance or extracting algorithms. This is the natural trade-off in between information spoil and privacy. Jiang and Clifton [14][15] targeted to 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 issue can be Reviewed such as a horizontally partitioned to data table in which each transaction are self by a different data provider. The output of their method is a classifier, but the result of our method is an anonymous mashup data that assist general data analysis or classification analysis [17].Secure multiparty computation (SMC) [23], [24] on the another side, allows sharing of the computed result (e.g., a classifier), yet entirely banned sharing of data. Final result 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] expand the work to focus the problem of high-dimensional anonymization of or the health science sector applying LKC-privacy [4]. All these activity 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 focuses the vertical integration problem.Trojer et al. [28] explain a service-oriented architecture for concealing K-anonymity in the privacy preserving data mashup hypothesis. Our paper is varying from this earlier activity [12], [13], [10], [11], [28] in two facets. First, our LKC-privacy model provides a stronger privacy assure than K-anonymity because K-anonymity does not recall the privacy threat due to attribute linkages, as mentioned in survey table second, our methodology can better preserve information utility in high-dimensional mashup data. High dimensionality is a captious obstacle for getting useful data mashup hence the combined data from more than one parties usually include several attribute. Our privacy model resolves the problem of high dimensionality.

3.PRPPOSED STATEMENT

We learn the privacy threats occur by data. The integrated table should be demonstrate both the following anonymity and information demand:

•**Anonymity Requirement**:

The integrated table has to fulfill k-anonymity A data table T fulfill k-anonymity if each combination of values on the QID is shared by at minimum k records in T, where the quasi-identifier (QID) is a set of attributes in T that can be to point out on an individual in T, and k is a user-defined threshold. K-anonymity can be fulfilled by centralizing domain values into higher level concepts. In addition, at any time in the methodology of generalization, no party should learn more brief information about the other party other than those in the final integrated table.

- **Information Requirement:**

The Generalized data must be as useful as possible to classification analysis. Generally, the privacy goal essential masking critical Data that is special plenty to find out singular, whereas the differentiate goal requires extracting trends and patterns that are general plenty 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 secrecy and information Demand, the data mashup application is an online web application. The user randomly explaining their requirement and the system is becoming to be efficient and countable to handle high volumes of data.

Privacy-preserving data mashup having multiple private tables T1, . . . , Tn, a joint anonymity demand {"QID1, k1", . . . , QIDp, kp"}, and To generalize T, a taxonomy tree is make up for each categorical attribute in UQIDj. For a numerical attribute in UQIDj , a taxonomy tree can be expand at dynamic, where each node shows an interval, and each non-leaf node has two child nodes show some optimal in two split of the parent interval. The algorithm re size a table T by a series of specializations starting from the top most general state in which each attribute has the top highest value of its taxonomy tree. A specialization, mostly written as v →child(v), where a child(v) demonstrate the set of child values of v, displace 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 issue of privacy-preserving data mashup is to efficiently generate a generalized Combined table T such that

1. T fulfill the joint anonymity demand,
2. contain as enough Data as possible for classification, and
3. Every party study nothing about the other party more special than what is in the final generalized

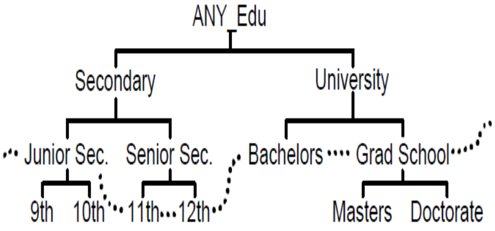
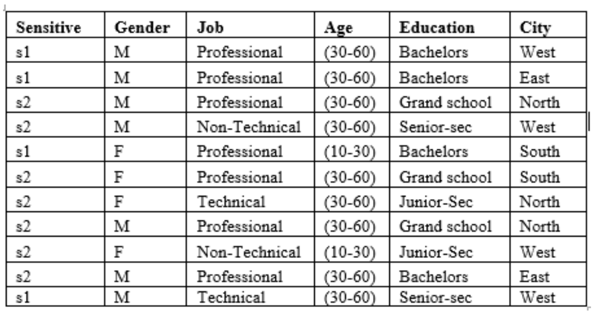


Fig. 1.Taxonomy Tree and QIDs.



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 Separately, and Attched the generalized tables to produce the integrated data. However, if there are global QIDs, global QIDs are neglect in this approach. Further generalizing the integrated table using global QIDs does not active because the requirement (3) is desecrated by the internal table that includes 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 neglect the opportunity of getting a better result by generalizing QIDg first, which leads to a sub-optimal solution. A finer strategy is generalizing shared attributes in the presence of both QIDi and QIDg. Similarly, the generalization of shared attributes will make change 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.

4. MATHEMATICAL MODEL

Consider n data originator Originator 1 ……….Originator n where every provider y own a private table T (UID,QIDy,Sy,Class) over identical group of records.UID and Class are same attributes through entire all data providers. QIDy is a set of quasi-identifying attributes and Sy is group of sensitive values owned by provider y.

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

These providers ready to open “minimal information” to create a mashup table T (by coordinated the UID) for arranging general data analysis or a joint classification analysis. The concept of minimal information is announced 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 effective admirer, i.e. there is one a polynomial time algorithm that can check if a given answer is right.

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

- A specialization is valid if the specialization results in a table fulfill the anonymity demand after the specialization.
- A specialization is use full if multiple class are involved in the records containing.

The verifier V gets two inputs,

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

One technique is computing Score, which count the efficiency of a specialization with respect to privacy elaboration and information protection.

The effect of a specialization v → child (v) can be checked by information gain, declare Info Gain(v), and anonymity loss, declare AnonyLoss(v), due to the specialization. Our selection method is to consider the specialization v that has the maximum information gain per unit anonymity loss. (1) We add 1 to AnonyLoss (v) to restrict division by zero.InfoGain (v): Let T[x] assign the set of records in T generalized to the value x. Let freq(T[x]; cls) denote the multiple records in T[x] having P the class cls.Note that Where c ε child (v).We have (2)Where I(T[x]) is the entropy of T[x] :I(T[x])= (3) spontaneously, I(T[x]) measures the mix of classes for the records in T[x], and Info Gain(v) is the deduction of the mix by specializing v.AnonyLoss(v): This is the average loss of anonymity by specializing v on all QIDj that include the attribute of v: AnonyLoss(v)= avg{A(} (4) where A(QIDj) and Av(QIDj) display the anonymity earlier and later specializing v. Note that AnonyLoss(v) not just assemble on the attribute of v; it assemble on all QIDj that include the attribute of v. Hence, avg {A (QIDj) Av (QIDj)} is the average loss of all QIDj that include the attribute of v.

5. PROPOSED ARCHITECTURE AND PROOTOCOL

We demonstrate a service-oriented architecture (SOA) that represent the communication paths of all involved parties, followed by a privacy-preserving protocol that can efficiently find out a suboptimal solution for the above described problem.

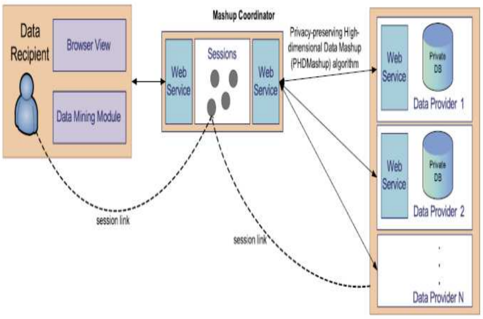


Figure2.Service-oriented architecture for privacy-preserving data mashup

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 get an information service request from the data recipient and set up connections with the data providers who can use their data to satisfy the request.
The objective of Phase I is to setup a common session context between the data recipient and the participating data providers. An Active context is successfully create and formed by proceeding through the iteration 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 dimensionality.

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 is valid(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 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 large parties At each step, the data providers help to do the same known specialization by communicating few count statistics information that fulfill requirement section 3.We represent the key steps: find the winner candidate (Lines 4-5), In Line 9, each party must communicate with all the other parties for finding the successor. Perform the winner specialization (Lines 7-9), Similarly, in Line 9, the party holding the winner candidate must briefing all the other parties and in Line 6, a party wait for instruction from the winner party.and update the score and status of candidates (Line 9).

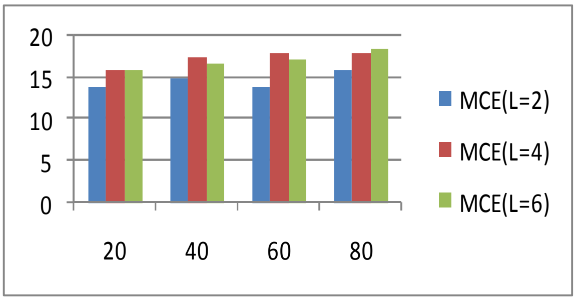
6. RESULT

We utilized the proposed PHDMashup in a distributed web service environment. Each data provider is using on an Intel Core2 Quad Q6600 2.4 GHz PC with 2 GB RAM connected to a LAN. To assets the benefit of data mashup for joint data analysis, Due to the privacy agreement, we unable 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 elaborate the performance of our proposed architecture and algorithm. The Adult data set has six numerical attributes, eight categorical and a binary Class attribute showing two income levels _50 K or >50 K. Table 3 explain every attribute. It includes 45,222 records after ejecting records with mismatch values. We model a 3-data provider scenario with three private tables TA ,TB, TC as follows: TA include 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 both categorical and numerical attributes are presents.

- ***Benefits of Mashup***

Depress classification error means finer data quality. We collect two kinds of classification errors from the testing set: Mashup Classification Error (MCE) is the error on the mashup data generate by our Centralized Anonymization algorithm. For large data providers. Source error (SE) is the error on separate 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 separate private table.

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



Threshold K
TABLE3: ADULT DATA SET
Fig. 3.Benefits of mashup(C=20%)

Fig. 3 show the MCE of the adversary's prior information L = 2, L = 4, and L = 6 with confidence threshold C =20% and an anonymity threshold K in range from 20 to 100. The benefit reducing as L improving because the most generalization is essential 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.

7. CONCLUSION

In this paper we studied how to combine and protect sensible data which deduce the privacy threat with the help of data mashup and recommend a service-oriented architecture for privacy-preserving data mashup so there the integrated data still keep the critical data for assisting general data search or a specific data mining activity, as like classification analysis

7. 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.

[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 byCharu c. aggarwal Philip s. yu kluwer academic publishers London

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