Secure Two Party High Dimensional Private Data Using Data Mash Up

Indhu Sridhar[1], Perm Jacob[2]

[1]M.E student, Sathyabama University, Chennai,
[2] Research Scholars, Sathyabama University, Chennai.

***Abstract*- The main scope of the project is to protect user's private sensitive information from other data providers using SOA (Service oriented architecture). Mash up is integrating different service providers to deliver high service to the customer. Data mash up is an application that aims at integrating data from various data providers depending on the user's request. Mash-up provide new platform to different data providers. Combining data from different source will reveal a person's sensitive information. Privacy Preserving Data Mash-up (PP Data Mash-up) algorithm is mainly used to hide the sensitive information. The algorithm securely integrate private data from other data providers where the integrated data still remains essential for supporting the general data**

Keywords: Sensitive information, mash-up

I. INTRODUCTION

Mash up is defined as a web application that gets content from various sources to create a single service. Mash up is generally considered as client application. Various types of mash up are available. We use consumer mash up which is aimed at general public to hide the user's sensitive information. By joining multiple data sets together will reveal the sensitive information to all the data providers.

i) The integrated or the mashed data could sharpen the identification of individual
ii) The mash up data from multiple sources contains many data attributes.
iii) Privacy Preserving Data Mash-Up is mainly for privacy threats by mash-up to securely integrate specific sensitive information from other data providers.

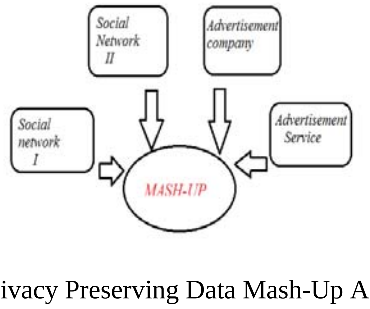

Fig 1. Privacy Preserving Data Mash-Up Algorithm

II. RELATED WORKS

Approaches have been proposed in literature for data mash up.

Anonymizing classification data for privacy preservation [1] reveals only explicit information. The demerit of this paper is the collection of anonymous properties to identify an individual.

Achieving K-anonymity privacy protection using generalization and suppression [2] states that only specific set of data is revealed to the third party and the demerit is shared information is revealed and user is identified.

L-diversity privacy beyond k-anonymity [3] reveals only useful information retrieved through data suppression in l-diversity. The demerit of this is sensitive information can be retrieved when there is some possibility in k-anonymity.

Protecting respondents identities in micro data release [4] states that table created using database, those data are matched with consecutive tables to get derived data. The demerit of this is all sensitive information is released after compression.

Workload aware anonymization [5] filters the specific information about the users using k-anonymity and l-diversity. And the main demerit of this is sensitive information is also revealed but, those are useless data.

III. EXISTING WORK

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 mash up. High dimensionality is a main obstacle for achieving effective data mash up because the integrated data from multiple parties usually contain many attributes. Tradition k-anonymity on high dimensional data will result in significant information loss.

IV. PROPOSED WORK

Mash up coordinator receives the entire information service request from the data recipient and setup a connection with the data provider who can share the data's to carry out the request. Mash up co-ordinator executes the privacy preserving algorithm to combine private data from multiple data providers and to convey the final mash up data to the data recipient. The proposed solution does not need the mash up co-ordinator to be a trusted party. Though the mash up co-ordinator is responsible for all the mash up service, over solution assures that the mash up co-ordinator does not gain more information than the final mash up data by protecting the data of the entire candidate.

V. Mashup

Mash up is defined as a technology that allows various service providers to flexibly integrate their expertise and to deliver high customizable service to all customers. Mash up coordinator will mash up the data provider and the third party. Two types of mash up are available. Web based mash-up and server based mash-up. The web based mash-up use the user's browser to combine the data and the server based mash up revise the data on the remote server and transmit the data to the user in a result format. The main businesses are adopting service oriented architecture to integrate data by making them as a web service. The web services provide an open standard protocol to provide a unified way for accessing information from platform.

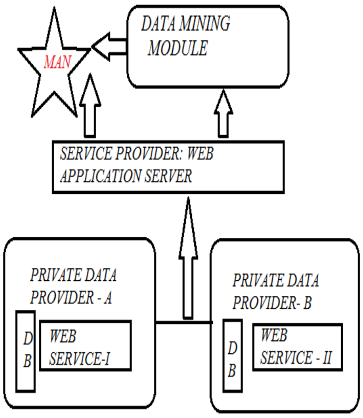

Fig 2. Mash-up algorithm

5.1 K-Anonymity

K-Anonymity is a main technique that is used. In this the user's information is restricted to a certain level so that the social network will provide only certain data's of the users. The sensitive information is protected. The adversary may get only user category but not all data. A specific table is considered T (Id, $D_1 \ldots D_{n,}$ class). ID defines record identifier and D is defined as a continuous attribute. Class column contain labels. The data provider wants to prevent against linking an individual to a record in T with subset of data called as Quasi Identifier (QID). A sensitive link occurs if few values of this QID are shared.

5.2 LKC-Privacy

A database T meets LKC privacy of and only if |T (qid)|>=K and Pr (S|T (qid)) <=c for any given attacker knowledge q, where |q|<= L

S – Sensitive attributes

K – Positive integers

Qid – adversary prior knowledge

T(qid) – group of records that contains qid

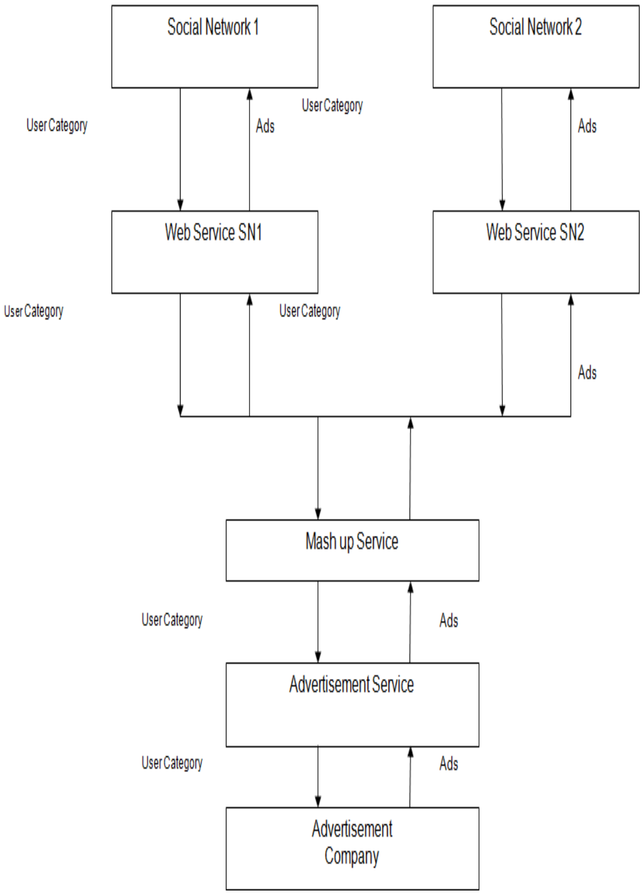

Fig 3. Mash up architecture

VI. Conclusion

Based on the mash up algorithm that is used in this mainly focuses on hiding the sensitive information from the third party agents. The data is being mashed up on the user end before it is sent to the third party. Only the common and general information is revealed to the third party agents. And the sensitive information is hidden and they are encrypted before it is sent to other agents. Through this all the sensitive information that we use to register on a social networking site is mashed up before it is sent to other party.

Reference

[1] *B.C.M Fung, K, Wan and P.S.Yu, "Anonymizing classification Data for Privacy Preservation,"* IEEE Trans .Knowledge and Data Eng., vol 19, no 5, 99 711-725, May 2007.

[2] *L.Sweney, "Achieving K-anonymity privacy protection using Generalization and Suppressions,"* Int'sJ.Uncertainty Fuzziness and knowledge based systems, vol 10 no 5 pp 571-288, 2002

[3] *A.Machanavaijhala, D.Kifer and M.Venkitasubramanian "l-Diversity Privacy Beyond k-Anonymity,"* ACM Trans Knowledge Discovery from Data, vol1, no 1, Mar 2007.

[4] *P.Samarati, "Protecting Respondents identities in micro data Release,"*IEEE Trans. Knowledge and Data Eng., vol.13, no.6, PP.1010-1027, Nov.2001

[5] *K.LeFevre, D.J.DeWitt, and R.Ramakrishnan, "Workload Aware anonymization"* Proc.12th ACM INT'l Conf. Knowledge Discovery and Data Mining (SIGKDD), Aug 2006.