Study of SOA and mashup Technique for Handling confidentiality of private data

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

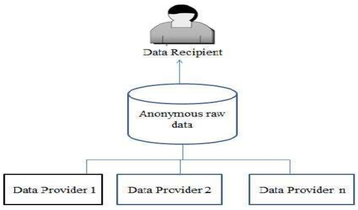
M. TechScholar Engineering, Department Computer Science [1,] Associate Prof Dept. of Computer Science Engineering [2,] , NRI-IIST Bhopal, M.P., State, India [1,] ,NRI-IIST Bhopal, M.P, State, India [2]
Email: dineshpatil83@gmail.com, umeshlilhore@gmail.com [2]

Abstract - Mash up Technique is based on web Technology that permits several service providers to flexibly integrate their intelligent and to deliver highly commercial and customize services to their consumer. Data mash up is a specific kind of mash up Technology that goal at combining data from several data providers as per user's requirement. However, integrating data from several sources brings about .Three challenges: simply adding multiple private data sets all together would disclose the critical information to the another data providers. The collected (mash up) data could potentially very sharpen the identification of individuals and, hence, reveal their person-specific critical information that was unavailable earlier to the mash up. The mash up data from several sources normally includes various data attributes. When targeting a ongoing privacy model, such as K-anonymity, the high-dimensional data would going from the issue called as the curse [3] of high dimensionality, resulting in unwanted data for next data analysis. In this paper, we are studying and formulate a privacy problem in a real-life mash up technique for the online advertising area in social community, and take initiative on a service-oriented architecture also using with a privacy-preserving .Data mash up program to call the said opportunity. Do practices on real-life data advise that our targeted architecture and algorithm is effective for one by one preserving both privacy and information module on the mash up data. To the best of as per our knowledge, it is the first work that combines high-dimensional data for mash up service.

Keywords—Mash up, Service-oriented Architecture, Privacy Model.

1. INTRODUCTION

A mashup is in web development, as like a web page, or web application, that taken content from multiple sources to established a single new service show in a single graphical interface. Here we explain one example; you could collect the detail such like addresses for location, and photographs identification of our library sub location with a Google map to create a latest map mashup. The term implies simple, fast combination, continuously using open application programming interfaces (API) and data sources to generate effective output that were not required the main reason for creating the unprocessed source data. The terminology mash up primarily comes from one type of music, where people smoothlessly combine music from one song with the instrumental track from another-thereby mashing them entirely to establish something new. First here discuss the Problem Statement and privacy goal is specified by the LKC privacy model on a combination of attributes called a implicit identifier, where each description on a implicit identifier is required to be given by some least number of records in the table. A generalization taxonomy tree is specified for each categorical attribute in a implicit identifier. We present a Top-Down Specialization (TDS) approach to generalize a table to satisfy the LKC privacy model requirement while preserving its usefulness to classification. Section II describes about Related Work for privacy preservation. Section III we study the privacy threats caused by data mashup and propose a service-oriented architecture (SOA) used for a privacy preserving data mashup system. In order to securely integrate private data from multiple participant. The generalized data will be useful and there is possible to data analysis. To formulate the privacy-preserving technique for high-dimensional data mashup issue. Section VI describes the Proposed Work for the privacy preservation Experimental results (Section V) on real-life data suggest that our method can effectively achieve a privacy requirement without compromising the information utility, and the with respect to Implicit Identifier (IID) = {Gender, Job, Age} will newly proposed architecture is scalable to large data sets. Data mining is the process of extracting useful, interesting, and previously was known information which is from huge data sets. The success of data mining relies on the availability of high quality data and useful and applicable information sharing. The collection of digital and electronics information by private firm, governments, corporations, and individuals has produced an environment that provides huge amount data mining and data analysis. Information Sharing has a long past in information technology. Traditional information sharing refers to exchanges of data between a data holder and a data recipient. Not only refer to the traditional one-to-one model, but also the most general models with several data holders and data recipients.Apply latest standardization of the information sharing protocols, as like extensible Markup Language (XML), protocol like Simple Object Access Protocol (SOAP) and language which describe web services as Web Services Description Language (WSDL) are catalysts for the recent development of information sharing technology.



and finally concludes this paper specialization leads to a violation of the anonymity requirement.

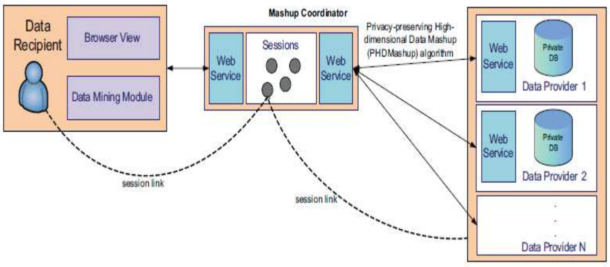
1 Additive mash up

The core specialty of a mashup is conjunction, visualization, and aggregation. It is crucial to do available data more effective, for individual and professional purpose. To be enabling to permanently right for data of another services mashups are mostly client side applications or hosted it to online.The concept mash up address to a newer community which is based on web. Applications made by hackers and several programmers mainly on a informer basis to mix minimum two different. Services from heterogeneous, and even capable, Web sites. A mash-up, for example, could overload travelling data from single source on the web over maps from Facebook, rediffmail,hotmail,tweeter, Google ,LinkedIn or any social Network.. The concept of mash up comes from the one kind of music it was hip hop practice of collect more than two songs or video.This capacity to combine and detect data and applications from several sources in one dynamic module is considered by various to show the ethics of the Web service standard (also address to as on-line computing).

2 Service-oriented architecture

Service oriented Architecture (SOA) .is pattern of a software design and software architecture design which is based on different module of software providing application working as services to other applications. This is called as Service-orientation. It is separated for any vendor, product or technology.this service is a own efficient unit of functionality such as accessing an online banking e-statement.. Services may be include by another inbuilt software applications to provide the whole working of a huge software application. it makes it simple for each and every computers which are in network to Enterprise. Each computer system can run an fixed number of services, and each service to be bind up in a such way that it assures these service can share information with any another service in the network without human intervention and Except the need to make changes to the running program itself. Service-oriented architecture (SOA) is an evolution of distributed computing it entirely based on the request and reply with two method design structure as synchronous and asynchronous. An application's business concept or individual working are modularized and presented as services for customer/End user applications. main advantages to these services is themes loosely coupled by default; i.e., the service interface is not dependent of the implementation. Application developers or system integrators can make applications by makings several services without knowledge of the respected services mention implementations. there given one example, a service can be utilized in .Net and J2EE, and the application giving the service can be on a distinct platform or language. The following are main characteristics of Service-oriented architecture its services are own-describing interfaces these are portable XML documents. Standardized Web Services Description Language (WSDL) used to show the services. its services mostly communicate with messages mainly defined through XML Schema (XSD). Communicate with consumers and providers or services mainly occur in different environments, with small or without knowledge about the provider. Messages between services can be visualize as main commercial documents doing in an organization.

3 Architecture



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

2. RELATED WORK

Information integration has been an most working field in database research. In this study w mainly assumes that entire information in each database can be freely shared Secure multiparty computation (SMC) , on the vise versa, permit publishing of the computed output (e.g., classifier), but completely prohibits sharing of data. An example is the secure multiparty computational work of classifiers. In contrast, the privacy-preserving datamashup issue studied in this paper allows data providers to publish data, not only the data mining results. In many applications, data sharing gives greater flexibility than result sharing hence the data recipients can do their essential analysis and data exploration. it target to the the sign of K-anonymity. Data file system and _-Argus system use generalization to cross K-anonymity. Then preserving classification information in K-anonymous data is learnt. Extend the work to address the issue of high-dimensional anonymization for the healthcare area using LKC-privacy. Entire these works consider a single data source; therefore, data mashup is not an issue. Adding entire private databases from more than one source and applying a single table anonymization method unable to assuarity of privacy if a QID span across more than one private table. Recently, th targeted to refer the horizontal integration trouble while this addresses the vertical integration trouble Propose a cryptographic approach and propose algorithm such as top-down specialization for securel combined two elevated differentiated distributed dat tables to attach a K-anonymous table, and furthe consider the participation of malicious parties. Thi Architecture for achieving K-anonymity in the privac preserving for data mashup scenario. The two mai aspects. First, our LKC-privacy model provides stronger privacy assurity than K-anonymity becaus K-anonymity does not call the privacy attacks due t by attribute linkages, .another, and our procedure ca better preserve information utility in high-dimensiona mashup data. High dimensionality is a required an important obstacle for getting impressive an applicable data mash up because the integrated dat from multiple participants normally include variou attributes. Reporting traditional K-anonymity o high-dimensional data will impact in remarkabl information loss. Our privacy model resolves the issu related about high dimensionality. This allegation i also subsidized by our experimental results. Develop cryptographic avenue to studied and understan classification norm from a huge number of dat providers while critical attributes are secured. Th problem can be viewed as a horizontally partitione data table in which every transaction is varied by different data provider. The result of their procedure i a classifier, but the result of our method is a anonymous mashup data that assist as usual dat analysis or by classification analysis. This demonstrat a privacy-preserving distributed data published for i line partitioned databases. The mashup model in thi paper can be viewed as a vertically partitioned dat table, which is very distinct from the model studied . secured communication mechanism that ability t cross-domain network appeal and client-side inte communication with the intention of protecting the mashup controller from malicious code through web services. In contrast, this aims to preserve the privacy and information service of the mash up data. Data mash up is one kind of special application which intent to integrate data from multiple data providers as per user Request. However, collecting data from multiple sources come with three challenges: Simply adding multiple private data sets all together would reveal the important information to the another data providers.[6] The collected (mash up) data could able to sharpen to find out individuals and, hence, reveal their person-specific critical information that was unavailable before the mash up. Its data from several sources generally include various data attributes.

3. PROPOSED WORK

A new privacy issue after collaboration of with the social community organization is find out and generalizes the industry's requirements [9] to formulate the privacy preserving of high volume (dimensional) data mash up issue. Service oriented architecture is used privacy-preserving data mash up as secure combine local data from more than one party. The knowledge about the privacy threats due to the data mash up and targeted to a service oriented architecture and a privacy preserving for data mash up algorithm [8] to carefully with keeping security combine person-specific important and sensitive data from multiple data providers, then the integrated data still give the essential information for helping general data exploration or particular data mining task.

We initial describe a service-oriented architecture (SOA) which show the communication paths of all participating parties; followed by a privacy-preserving high-dimensional data mashup algorithm that cans efficiently identify suboptimal solution. SOA is an architectural structural framework for create, creat, then develop and integrate various information systems with an exact message-driven communication model. Which Follow the SOA design principles, the final system has various preferable properties including ability and loosely coupling. Ability address to the efficiency of permit platform-independent (portable) design of the system unit based on general grasping of service component interfaces. Which Loosely coupling refers to the eligibility of reducing dependencies entire system unit and, hence, increasing the overall, scalability, flexibility and fault permissive of a system. In the mashup system explaining in this paper, data sources can be runtime composed to serve new mashup requests depending on the data analysis work and privacy demand. SOA with the efficiency of ability and loosely coupling has become a inbuilt choice to aggress the heterogeneity of different potential data providers.

Advantage

•Keep a privacy assurance.
•End Use Detail will be stored in its own database.

•Information utility preserves in high-dimensional mash up data.

•securing the mash up controller from malicious code which come various web services.

Disadvantage

•Database of user will be stored in web service provider.

•There is no secrecy and data's are not unprotected.

• K-anonymity cannot find out privacy attack because of linkage in data attribute

Parameters

The technologically driven world in which we live for improve human interaction with computer based, particularly with internet-based system that are used to fulfil a large variety of activity with the goal of assist the user in achieving targeted. The data mashup process can be partitioned into two phases. In Phase I, the mashup handlers get an information service request from the data receiver and create connections with the data providers who can contribute their own data to fulfill the demand. In Phase II, the mashup coordinator executes the algorithm of privacy-preserving to collect the private data from more than one data providers and to pass the final mashup data to the data receiver. Our proposed solution does not necessary the mash up coordinator to be a trusted party. Though the mash up coordinator Handel and carryout the entire mash up service, our solution assures that the mashup coordinator does not obtain more information than the final mashup data, thereby securing the data privacy of each participant. The mashup coordinator can be any one whose of the data providers or an independent party. This creates our architecture practical so a trusted party is unavailable in real-life mash up scenarios.

Phase I: Session Establishment

The objective of Phase I is to formed a common session context between the data receiver and the participated data providers. Usable discourse is successfully created by proceeding through the steps of data deliver authentication, participating data provider’s find out, meet context initialization, and common requirements negotiation. Authenticate data recipient. The mashup coordinators initially authenticates a data recipient to the requested service, session token created for the current receiver interaction, and then find out the data providers available by the data recipient. Different and data providers having public and are available by any data receiver. Finding data providers whose contribution is more. Next, the mash up coordinator fires the queries to the data schema of the available data providers to find them that can share data for the called service. To assist more effective queries, the mash up coordinator could executive data schema from the data providers or the data providers could update their data schema periodically one by one (i.e. the push model).start session context. Next, the mash up coordinator acknowledge entire participating data providers with the session identifier. Entire likely data providers share a common session context that show a one state presentation of information related to a specific execution of the privacy preserving mashup algorithm called PHDMashup. An established session context contains several attributes to identify aPHDMashup process, including the data recipient’s address; the data provider’s details such like addresses and certificates. Token of an authentication which contains the data recipient’s certificate. And a unique session identifier which use and apply an end-point reference (EPR) composed of the service address, aPHDMashup process identifier and runtime update and status of information about the executed PHDMashup algorithm. Bargain privacy and information requirements. The mash up coordinator is liable to communicate the bargaining of privacy and information demand on the data providers and the data recipient. Specifically, this step involves negotiating cost, LKC-privacy demand, important information, and required information quality. For example, in the case of classification analysis the quality of information can be determined by classification error on few testing data.

Phase II: Privacy-Preserving High-Dimensional Data Mash up

The objective of Phase II is to integrate the high-dimensional data which is from multiple data providers such that the resulting mash up data fulfills a given LKC-privacy requirement and preserves as much information as possible for the mentioned information demand. Requirements 1 and 2 specify the properties of the final mash up data. Requirement states that not allow to data provider should grasp more brief information than the last mashup data occur in the process of integration. To fulfill demand , we target a top-down specialization is known as Privacy-preserving High-dimensional Data Mash up (PHDMashup).

5 SCOPE

The correct demand to be utilized is listed below. This is an abstraction of the requirements sign-off document. Having demand in the project quality plan helps validation perform by the quality assurance team too correctly by this method quality confidencefunction knows what exactly to leave out from the scope. Testing the requirements that are not in the scope maybe a exploit for the services provider.

5 FUTURE RESEARCH

5.1 Privacy Measure:

Mash up organizer declares entire participating data providers with the session identifier. All like data providers share a common session context that show a specific and complete presentation of information related to a specific operation of the privacy-preserving mash up called PHDMashup [6]. An created session context several attributes to identify a PHDMashup process, containing the data receiver address; the data providers 'certificates and addresses; the token of an authentication in that consist the data recipient's certificate and a unique session identifier that uses an end-point mention composed of the service, a PHDMashup technique identifier and run time progress information about the operated.

5.2 Anonymous Mash up data:

The mash up data taken from several data providers usually contains many attributes. When trying apply convential privacy models [1] for high dimensional data in output the significant information loss.. if the number of attributes increases then more generalization is required for achieving k anonymity[5] even if k is little bit then final data useless for further analysis.

5.3 Raw Data Method:

A specific kind of mash up application which objective at collecting data from several data providers as per service request from a user [10]. An information service request could be a general count statistic activity or a refined data mining task as like classification analysis. as receiving a request of service then the data mash up web application dynamically find out the data providers taken information from them through their web service interface, and then after combine the collected information to fulfill the service request. next to computation and visualization can be handled at the user's site or on the web application server. This is very different from a conventional web portal that simply separates a web page or a website into separate sections for displaying information from distinct sources.Privacy-Preserving High-Dimensional Data Mash up: Integrating the high dimensional data [2]which come from multiple data provider is the main objective of phase II such that the last mash up data fulfill given requirement and preserves more information as possible for the mentioned information requirement. Recall that specifies three requirements. Requirements specify the properties of the resulting mash up data. Requirement states that no data provider should learn more detailed information than the final mash up data during the process of integration. To satisfy requirement we are proposing a top- down specialization approach known as privacy preserving high dimensional data mash up.

6 CONCLUSION

We are Researched and reviewed previous works on most challenging publishing scenarios and Intention , consisting multiple release publishing, sequential release publishing, continuous data publishing, and collaborative data publishing. Integrate and secrecy private Critical data is complex. with the help of data mashup and SOA architecture to privacy preserving data mashup. We can build model which only Extract only needful Information not whole.

REFERENCES

[1] 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.

[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. Know. Data Eng. 13(6): 1010-1027 (2001).

[5] 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

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

[7] [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] [8]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.

[9] [9] P. Jurczyk and L. Xiong, "Privacy-Preserving Data Publishing for Horizontally Partitioned Databases," Proc. 17th ACM Conf. Information and Knowledge Management, Oct. 2008.

[10] [10] 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..