Implementation of Data Privacy & Query Optimization in Data Providing Service Composition

Rashmi Kulkarni[1], Amit Chougule[2]

**[1] Assistant Professor, D. Y. Patil College of Engineering & Technology,
Kolhapur, Maharashtra, India**

**[2] Professor, Bharati Vidyapeeth's College of Engineering,
Kolhapur, Maharashtra, India**

Abstract

Data providing services allow query like access to organization's data via web services. The invocation of DP service results in the execution of a query over data sources. In most cases user's queries require the composition of several services. The RDF based query rewriting algorithm is used for DP services composition which has limitation that, it is unable to provide security for a confidential data which is represented in RDF graph. Hence the proposed system will use RDF Encryption in which confidential data in a RDF-graph will be encrypted for a set of recipients while all non-sensitive data remain publicly readable. Also the proposed system will use query optimization technique which will reduce the execution cost.

Keywords*: RDF, RDFS, SPARQL.*

1 Introduction

Modern enterprises are increasingly using service oriented paradigm to provide interoperable and programmatic interaction with internal system. Such interactions are generally performed via Data Providing services (DP). DP service allows query like access to Organization's Data sources. The invocation of DP service results in execution of query over data sources schema.

Such invocation has no effect on the state of world, e. g. Pharmaceutical DP service may return generic equivalent of given brand. In most cases user's queries require the invocation of several services. e. g. let us consider a query "what are the tests performed in ABC lab by patients who have been administered Glucophage in XYZ hospital". Let us assume that ABC Lab and XWZ hospital provide two DP services SABC and SXYZ, respectively SABC returns the tests performed by a given patient in ABC Lab and SXWZ returns the list of patients that have been administered a given drug in XWZ hospital. The execution of the above-mentioned query involves the composition of SABC and SXYZ services.

Current system provides a way for composing DP services. But it fails to address data privacy issue while composing a DP service. The proposed approach is to provide data security using RDF encryption technique with SPARQL queries and to use query optimization to speed up execution of the query. The organization of this paper is as follows:

Section 2 defines the terminologies used in this paper,
Section 3 focuses on related work,
Section 4 explains need of data privacy in Semantic web,
Section 5 gives idea about Proposed system and
Section 6 concludes the paper.

2 Preliminaries

2.1 RDF

Resource Description Framework (RDF) is a language that can represent information, particularly metadata like author or title, about resources in the Web. RDF is intended for the Semantic Web [8]. It has to be machine readable and understandable. RDF uses Uniform Resource Identifiers (URIs) to identify things. Like this, RDF can represent resources, properties and values as a graph. Individuals, kinds of things, properties of those things and values of those properties are identified by URIs.

2.2 RDF Schema (RDFS)

It is a simple modeling language on top of RDF which includes classes, is-a relationship between classes and between properties, and domain/range restrictions for properties. RDF and RDF Schema are written in XML syntax.

2.3 SPARQL

SPARQL is standardized, supported and recommended by the World Wide Web Consortium (W3C). RDF data is represented as a graph; SPARQL is a graph querying language and can be used on one or more RDF files in the memory.

2.4 Ontology

Ontology stands as the most important concept in Semantic Web, which can be defined as a collection of key concepts and their inter relationship providing an abstract view of an application domain. The Ontology enables both user and system to communicate with each other by the shared and common understanding of a domain. For the web, ontology is about the exact description of web information or web resources and the relationship between them.

3 Related Work

3.1 Query Rewriting Approach for Web Service Composition

Semantic Web services are usually modeled with the de facto Standard for service description OWL-S2 [7]. In particular, OWLS's Service Profile permits the modeling of the service's input, output, its mandated preconditions and the produced effects of invoking the service. It also allows the categorization of services according to their functionality in a domain using the Service Category class in conjunction with some categorization schemes available within an application domain. DP services on the other hand are simply concerned with retrieving the appropriate output data given a specific input. They do not provide any functionality, beyond retrieval, and have no external effects. So some researchers have proposed RDF based Query Rewriting algorithm [2] for composition of DP service in which the services are modeled as RDF views and RDF views are enriched with RDFS semantic constraints and used to annotate WSDL descriptions.

3.2 Integrating Heterogeneous Data Source Using Ontology

Integrating data from multiple heterogeneous sources [9] entail dealing with different data models, schemas and query languages**.** The burgeoning Semantic Web has provided several methods for integration of heterogeneous data sources.
Being “explicit specification of a conceptualization” ontology is considered as a possible solution to represent the content of heterogeneous data sources. RDF is a general proposition language for description of ontology. SPARQL, a query language for RDF, can join data from different databases, as well as documents, inference engines, or anything else that might express its knowledge as a directed labeled graph.

4 Need of Data Privacy in Semantic Web

Giving information a well-defined meaning is on one hand the basis for intelligent applications in an emerging Semantic Web, but on the other hand can have profound consequences when considering privacy, security, and intellectual property rights issues. In the Semantic Web vision agents automatically gather and merge semantically annotated data, infer new data and re-use the data in different contexts .However seemingly harmless pieces of data could reveal a lot of information when combined with others. This section explains different approaches to implement Data privacy in semantic web application.

4.1 Controlling Access to RDF Graphs

The number of applications that publish and exchange possibly sensitive RDF data continuously increases in a large number of domains ranging from bioinformatics to e-government. Unfortunately, the potential of these efforts and the realization of the Future Internet is undermined by the lack of an _effective mechanism for controlling access to such data. In light of the sensitive nature of the information available, the issue of securing RDF content and ensuring the selective exposure of information to deterrent classes of users depending on their access privileges is becoming all the more important. The building blocks of an access control system are the specification language[3] that allows the expression of access control permissions and policies, and the en-forcemeat mechanism, responsible for applying the latter to the data, in effect denying access to non-accessible data.

4.2 Encryption of RDF – Graphs

Partial RDF encryption (PRE)[1] which allows for fine-grained encryption of arbitrary fragments of an RDF-graph without creating additional resources. Both encrypted data and plaintext data are represented in a single RDF-compliant model together with the metadata describing the encryption parameters. PRE uses the XML-Encryption and

XML-Signature standards to represent the encryption metadata.

Encryption policies for RDF-graphs define which fragments to encrypt and how to encrypt them. The PRE Policy Language (PRE-PL) uses a graph pattern based approach that allows for dynamic selection of encryption fragments. PREPL uses the RDQL[5][4] query language. The result of a query can be interpreted as a set of fragments which are instances of the same ’category’ defined by the search pattern. Each category is encrypted in the same way (the same keys, algorithms, etc.)

4.3 Comments

From above literature survey we can state that

(i) DP services are modeled using RDF views which will capture the semantic relationship between input and output from mediated ontology. There is no any mechanism to protect sensitive data in RDF graph while composing DP services.

(ii) Access rights can be specified for controlling data access and to secure communication channel when the data is transferred but this approach needs a trustworthy infrastructure for specifying and controlling data access.

(iii) RDF encryption techniques can be used to protect sensitive data in the RDF graph. RDQL query language is used to select encryption fragments but RDQL language does not support closure and optional pattern matching concepts.

(iv) SPARQL supports querying of multiple RDF graphs. However, the current standard does not provide transparent query federation, which makes query formulation hard and lengthy.

5 Proposed Work

To overcome above discussed drawbacks, we propose a new system

(i) To implement Data Privacy in the composition of DP services using RDF encryption with SPARQL queries

and

(ii)To speed up query execution in DP Service Composition using DARQ [6].

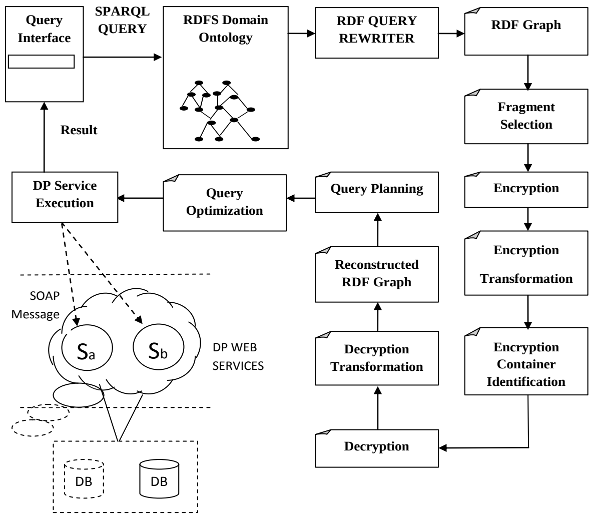

Fig. 1 Proposed System

The Proposed system will contain following modules.

5.1 Designing a web based Query Interface

Web-based query interface will be used to specify input queries (RDF queries) over the mediated ontology.

5.2 Modeling DP service as a RDF View

This module will model DP Service as RDF views over domain Ontology. Input / output relationships will be represented based on concepts and relations that are semantically defined in a mediated ontology.

5.3 Composing DP Services by Query Rewriting

Given a Query and a set of services represented by their corresponding views a rewriting of –query using the services will be constructed by a composition of services whose RDF Graph’s union covers the RDF graph of the query.

5.4 Applying RDF Encryption to the RDF Graphs:

Union of RDF graph is given as an input to the RDF Encryption. It will have following steps.

i. Fragment Selection: The first step will be the selection of the RDF fragments to be encrypted. RDF fragments can either be subjects, predicates, objects, or triples.

ii. Encryption: In this step, each encryption fragment is serialized and encrypted. The result of this step is a Encryption Container containing both, encrypted Data and encryption metadata.

iii. Encryption Transformations: All encryption fragments are replaced by their corresponding encryption containers. The result will be a single self describing RDF compliant graph containing three different kinds of components:
   (1) encrypted data, (2) encryption metadata and (3) plaintext fragments.

iv. Encryption Container Identification: Encryption containers and encryption Metadata are identified and extracted. This will be done by using SPARQL query language.

5.5 Decryption

In this step, the encryption containers are decrypted according to the parameters specified in the encryption metadata. If a receiver does not have an appropriate decryption key, the decryption process will fail.

5.6 Decryption Transformations

The last step will be the re-construction of the RDF-graph by replacing the encryption containers with the corresponding decrypted values. Graph transformations have to be performed which are inverse to the encryption transformations in step three. If a recipient has the keys to decrypt all encryption containers, then the re-constructed RDF graph is identical to original RDF-graph.

5.7 Query Planning

The original graph obtained from Decryption Transformations (which is parsed format of SPARQL query) will be used for Query Planning in which the query engine will decompose the query and will build multiple sub-queries according to the information in the service descriptions.

5.8 Query Optimization and Execution

The query optimizer will take the sub-queries and will build an optimized query execution plan which will be then executed and results will be shown on query interface

6 Conclusion

It is observed that DP service composition lacks in providing security for private data which is represented in RDF graph so RDF encryption technique will be used to protect the data in which SPARQL queries will be used to select encryption fragments from RDF graph, which provides the features like closures and optional pattern matching.

SPARQL load all RDF graphs mentioned in a query to the local machine. This usually incurs a large overhead in network traffic, and sometimes is simply impossible for technical or legal reasons. To overcome these problems DARQ, an engine for federated SPARQL queries will be used for Query Optimization in DP service Composition.

References

[1] Mark Giereth, “On Partial Encryption of RDF-Graphs”, ISWC 2005, LNCS 3729, pp. 308–322, 2005._c Springer-Verlag Berlin Heidelberg 2005.

[2] Mahmud Barhamgi, Djamal Benslimane, and Brahim Medjahed, “Query Rewriting Approach for Web Service Composition.” , IEEE Transactions On Services Computing, VOL. 3, NO. 3, July-September 2010.

[3] Giorgos Flouris1, Irini Fundulaki1, Maria Michou and Grigoris Antoniou,Institute of Computer Science, FORTH, Greece Computer Science Department, University of Crete, Greece. “Controlling Access to RDF Graph”, http://data.gov.uk/, http://www.data.gov/.

[4] Cristian P´erez de Laborda and Stefan Conrad, “Querying Relational Databases with RDQL”,Berliner XML Tage 2005, Berlin, Germany, September 2006.

[5] Andrea Liechti Zelgweg 40, “Search Result Ranking for a Reputation Analysis Prototype”, andrea.liechti@unifr.ch — Student-Number: 08-213-340, February 26th, 2012 ,Information Systems Research Group.

[6] B. Quilitz and U. Leser, “Querying Distributed RDF Data Sources with SPARQL,” Proc. Fifth European Semantic Web Conf. (ESWC ’08), The Semantic Web: Research and Applications, vol. 5021/2008, pp. 524-538, 2008.

[7] D.Calvanese, G.D. Giacomo, M. Lenzerini, M. Mecella, and F.Patrizi, “Automatic Service Composition and Synthesis: theRoman Model,” IEEE Data Eng. Bull., vol. 31, no. 3, pp. 18-22, 2008.

[8] Michael Grobe “RDF, Jena, SparQL and the “Semantic Web”, Indiana University Indianapolis, Indiana USA1.317.278.6891.

[9] Jinpeng Wang, Jianjiang Lu, Yafei Zhang, Zhuang Miao and Bo Zhou “Integrating Heterogeneous Data Source Using Ontology”, Institute of Command Automation, PLA University of Science and Technology, Nanjing, China , Journal of software, vol. 4, no. 8, october 2009.

Rashmi Kulkarni is persueing M. E. (Computer Science Engineering) at D. Y. Patil College of Engineering, Kolhapur, State – Maharashtra, County – India and is Assistant Professor in Department of Information Technology at the D. Y. Patil College of Engineering, Kolhapur, State – Maharashtra, County – India. She received B. E. (Information Technology) with Distinction at Shivaji University, Kolhapur, State – Maharashtra, County – India, in 2006

Amit Chougule is M. Tech. (Computer Science Engineering) from Shivaji University, Kolhapur, State – Maharashtra, County – India and is Professor in Department of Computer Engineering at Bharati Vidyapeeth’s College of Engineering, Kolhapur, State – Maharashtra, County – India.