Review on Authorization control for a semantic data repository and database management system

Neelesh Devre
Department of CTA, TIT College
RGPV, Bhopal, M.P. India
neelesh_devre@rediffmail.com

Abstract— Due to the increasing amount of data available online, the World Wide Web has becoming one of the most valuable resources for information retrievals and knowledge discoveries. The Semantic Web is well recognized as an effective infrastructure to enhance visibility of knowledge on the Web. The core of the Semantic Web is ontology, which is used to explicitly represent our conceptualizations. The Authorization control has become an important field of research for many software systems. It is related to database security which is concerned with preventing any unauthorized user to have access on any confidential information. Access control is an essential operation in standard information systems to prevent unauthorized access and use of information from the system. Due to various data sensitivities various authorization and access control mechanisms are needed in the area of data base security. Semantics of data should be considered in order to specify effective authorization control techniques. In this paper we review the various methodologies for authorization control of a semantic data repository.

Keyword: Authorization and access control, Semantic data repository, RDF, OWL, XML.

I. INTRODUCTION

The current Web is largely built on HTML[1]. Problem of the current web architectures is the Web systems are not designed to “understand” the Web content on their own. Nowadays, the World Wide Web has becoming one of the most comprehensive information resources. Web mining technologies are the right solutions for knowledge discovery on the Web. The Semantic Web is designed to solve this problem, by enriching web content with markup data. This markup data means to add more structural information to the semi-structured information in HTML page. This markup data gain benefits in machines understandability. Therefore it can enhance agent application to process web content. There is also close relationship between ontology and semantic web as ontology is the key elements for building up of semantic web content.

The Semantic Web [2] is designed to let users make explicit statements about any resource, and maintain that data themselves in an open and distributed manner. To enhance efficiency of information retrieval, several web mining techniques have been proposed including methods deriving from data analysis and conceptual analysis. With the ability of intelligent analyses, it can help people acquire appropriate information and discovery the latent semantic knowledge effectively. Nowadays, semantic web and ontology have shown their usefulness in application areas such as intelligent information integration, information brokering and Natural Language processing. The core technique of Semantic web mining is ontology. In computer science, ontology represents a set of precisely defined terms about a specific domain and accepted by this domain’s community. Ontology is an explicit specification of a conceptualization.

In recent years, the use of database security pays an important role in all areas which increase the security in storage and transmission of confidential information and therefore research in this area is growing rapidly. Security breaches can be classified as unauthorized data observation; incorrect data modification and unavailability of data. The extremely rapid development of the internet brings more and more attention to the database security techniques, most conventional authorization control algorithms are based on access control policies of organization.

Authorization plays an important role in database management system and semantic data repository. Organizations are needed to protect knowledge against unauthorized users. In this paper we are focusing on authorization and access control on semantic data repository and database management system.

Semantic data repository is used to manage semantic data model but it cannot deal with access rights. While in traditional databases semantic data is protected by granting access rights to authorized user only. In recent years same attention is needed for semantic data repository. Granting authorization privileges to semantic repository is used to prevent knowledge leakage from knowledge bases.

II. BACKGROUND

A. Extensible Markup Language

The eXtensible Markup Language (XML)[2] has become a standard language for data representation and exchange. XML is a Standard, flexible *syntax* for data exchanging Regular, structured data. With the continuous growth in XML data sources, the ability to manage collections of XML documents and discover knowledge from them for decision support becomes increasingly important. XML allows the representation of semi-structured and hierarchal data containing not only the values of individual items but also the relationships between data items. Element tags and their nesting therein dictate the structure of an XML document. Due to the inherent flexibility of XML, in both structure and semantics, discovering knowledge from XML data is faced with new challenges as well as benefits.

B. RDF: Resource Description Framework (RDF)

The RDF[3] is a simple Meta model for defining and exchanging information on the semantic web. In the semantic Web vision, the Resource Description Framework (RDF) data model provides a framework to capture the meaning of an entity[4] (or resource) by specifying how it relates to other entities (or classes of resources). Each of these relationships between entities is what we call a "semantic association" and users can formulate queries to find the semantic association(s).

C. Ontology

An ontology[5] is an explicit specification of a conceptualization. It contains four implications: conceptualization, explicit, formal and share. Ontology is used to describe the common and shared concepts and their relations in a specific-domain. It makes these concepts and their relations have common approbatory, explicit and exclusive definition in the shared area. As sharing conceptualization of knowledge presentation, ontology has been used widely in many domains, such as knowledge engineering, knowledge management, intelligent information

integration, information retrieval, semantic web and digital library etc. Ontology together with a set of individual instances of classes constitutes knowledge bases. Ontologies will play a pivotal role in the Semantic Web by providing a source of shared and precisely defined terms that can be used in meta-data.

D. Authorization control

Authorization[6] is finding out if the person, once identified, is permitted to have the resource. This is usually determined by finding out if that person is a part of a particular group, if that person has paid admission, or has a particular level of security clearance. The need of authorization control: The key aim of authorization access control mode[7] is to provide security system related to storage of confidential data. Semantic databases are gaining attention in the areas of industry, health care, content management and life sciences. A repository is a heart of storage and management systems extracting facts and information.

E. Semantic data repository

Semantic repository[8] is an engine similar to database management systems (DBMS) that permits the storage, querying and handling of structured data. In addition, semantic repository uses ontology's as semantic schemata to automatically reason about the queried data. Semantic repositories make use of generic and flexible physical data models, such as graphs. This permits them to quickly read and implement new metadata schemata or ontology. As a result, semantic repositories provide better incorporation of assorted data as well as more analytical power. However, these kinds of repositories are still in the early stages of their development. Databases are used in various kinds of applications such as surveillances, record keepings in medical fields, military fields, storage of confidential documents in defense systems, criminal related information's in investigation fields, etc. These semantic databases are most vulnerable to unauthorized accesses by eavesdroppers with an intention of stealing the confidential data. Hence there is the need of restricting the very access to the database by unauthorized users along with providing, the security to the inner contents of the database. This dual approach can encapsulate the secret information in the databases under two protecting covers, that is, restricting access as well as securing the database contents. Therefore a system which makes use of some notifications can be employed which alerts the authorized users of the databases in case of any unauthorized access activity being performed by an unauthorized user with a view of getting an entry into the database and even if someone succeeds to get an entry into database, the contents are not easy to find off as they are secured by some kind of encryption mechanism. Semantic data repositories expand the functionality of ordinary storage engine. The semantic data repository is a term that can defined database Management Systems that can be used to store, query and manage data structured according to the Resource Description Framework) standard. Compared to RDBMS such systems use flexible ontological schemata where data processing is done by an inference-engine according to a well-defined semantics. Fine grained access rights are not properly implemented in semantic data repositories. In semantic data repositories a user can infer any confidential information by applying logics and reasoning because inference policy[8] engine is used in such repositories.

III. NEED OF PROVIDING SECURITY TO DATABASE AND REPOSITORIES

Databases are used in various kinds of applications such as Storing data, record keepings in several fields, military fields, storage of confidential details in many systems, criminal related information in investigation fields, etc. These databases and storage repositories are most vulnerable to unauthorized accesses by attackers with an intention of stealing the confidential data. Hence there is the need of restricting the access to the database and semantic data repositories by unauthorized users along with providing, the security to the inner contents. This approach can encapsulate the secret information in the databases under two protecting covers, that is, restricting access as well as securing the contents. Therefore a system which makes use of some notifications can be employed which alerts the authorized users in case of any unauthorized access activity being performed by an unauthorized user with a review of getting an entry into the database or repository and even if someone succeeds to get an entry into database, the contents are not easy to get as they are secured by some kind of encryption mechanism. The databases can be made more secure by providing security at two different levels, that is, access level security[8] and content level security.

IV. LITERATURE SURVEY

[8] proposed a semantic reasoner authorization model conceived to secure access to semantic repositories. The model is able to restrict access to confidential knowledge that is represented using ontologies and can regulate access to both the ontology's concepts and their individuals. Therefore, the model fully supports content based access control. From a functional perspective, the model consists of two main components: TBox access control for the construction of the TBox family and ABox label-based access control for facts in the domain knowledge. The model also supports a flexible mechanism for propagating authorizations through a hierarchy concept tree. Experiments and evaluation shows that the complexity of the authorization model does not affect reasoning results or modularization. Introduce two-level access control paradigms (TBox and ABox) to provide highly secure operations for safeguarding semantic data repositories.

Provide TBox access control for the construction of a TBox family that regulates access to the concept individual level by designing access on a concept level. Propagate authorization-based construction of concept taxonomies for the family TBox. Provide ABox label-based access control for facts in

the domain knowledge. Create two level authorization controls for securing semantic knowledge TBox and ABox to enable the model to manage privileges of the users and roles on TBox objects exclusively, so, the policy-based storage optimization can achieve efficiency in the model's performance. This authorization mechanism fully supports content based access control, so that the authorization requirements are established not only for the model's concepts in the TBox, but also for their individuals in the ABox.

Other research has been conducted on numerous security aspects, consisting of authorization regarding semantic models. Dietzold and Auer [9] discussed the necessary entities for an access control model that targets an RDF triple store, generically, and developed a very basic model from them. The model is based on the utilization of filters that select the triples that a user can access according to their credentials; the original request is then executed on the filtered triples. Semantic access control has been integrated with a mediator; in other words, a software layer that offers a uniform interface to a set of heterogeneous data sources. An ontology is utilized to map the database schemas that need to interoperate, while a table records the communication among the roles of the various databases (role-based access control is assumed). Utilizing this table, the system ascertains whether a user should be given access to objects in a specified database, depending on the user's query.

Traditional discretionary access control, without data alteration operators, applied directly on ontologies can result in revealing unintended information because ontologies contain meta-information about objects. As an alternative author provide a constraint logic programming based policy language that can extract, remove or desensitize sensitive concepts in ontologies prior to requested disclosures. This policies are stratified Horn clauses with constructive negation, and constraint system uses a finitary system of ZF sets developed by Dovier et al. - and consequently, admits a three valued Kripke-Kleene semantics. Consequently, it is suitable for safeguarding meta-information stored on the Semantic web using OWL.

The main contribution of this paper is a policy-based disclosure control framework for safe sharing of sensitive ontologies. Authors approach can prevent disclosing sensitive portions of an ontology, selectively hide names of concepts and/or relationships while disclosing the overall structure of the ontology, and replace them with desensitized names - thereby allowing the access controller to spread deceptive information or cover stories to requestors with known attributes. The framework is based on constraint logic programming (CLP) based scheme with set constraints of a computable set theory, referred to as Set. This system-defined predicates contain all the requisite logic and as a result, user defined policy rules are reduced to simple, non-recursive predicate definitions. However, this policy language can be easily extended to give greater control to an advanced user.

Qin and Atluri [6] proposed a control model consisting of concept-level access that supports propagation based on the relationships among concepts for the purpose of regulating the access to data by individuals. A policy language based on constraint logic programming principles has been adopted. This language is utilized to define a framework for controlling the disclosure of sensitive portions of an ontology, and for selectively hiding names of concepts and/or relationships and replacing them with desensitized names. Kagal et al. [9] proposed policy languages similar to Rei based on Semantic Web languages like RDF and DAML+OIL and developed a framework, Rein, based on Rei. Finin et al. [8] proposed the use of the OWL language as the formalization of the RBAC model. They provide two ways to formalize an RBAC role, as a class or as an attribute.

Qin and Atluri [6] and Javanmardi et al. [12] consider the implicit authorization propagation and authorization conflict problem for various semantic relations in an ontology. That is, besides the subClassOf and subPropertyOf relationships, they consider the equivalence relationship between two concepts, the partial relationship between whole concepts and partial concepts (for example, intersectionOf and unionOf in OWL), the non-inferable relationship (for example, disjointWith and differentFrom in OWL), etc. However, their access control policy is not based on the RDF triple. Hence, their methods are not incorporated with RDF and OWL specification, and especially RDF inference.

Another method, proposed by Chen and Stuckenschmidt [10], enforces access restrictions by means of query rewriting. This approach is proposed as a suitable way of enforcing access restrictions in the context of SPARQL queries, while the TBox is assumed to be completely public. Similar approaches also make it possible to hide TBox parts, or to define not the restrictions but the permissions by a query. The idea is to automatically add filter conditions to the query that suppress those answers that the user is not supposed. This way of rewriting the query based on the access restrictions of the individual users effectively prevents the system from giving away restricted knowledge. However, it comes with a problem: It hides more knowledge than necessary. Also, the TBox is not assumed to be completely public because an access control policy for an ontology should consider the propagation based on the semantic relationship among concepts (TBox family). Security concerns over ontologies can be violated if access control to concepts is considered separately outside of the concepts' relationship.

Kaushik et al. [9] introduces an access control model for the fine-grained information disclosure of an RDF web document. The main point of their study is to introduce a formal framework to provide disclosure control over parts of an ontology. In addition, they introduce applying several methods of information hiding to RDF data, e.g., removing a specific subtree in an ontology tree or renaming a disallowed class or property according to an authorization. However they do not consider the disclosure problem for highly sensitive data by a prohibited inference. In fact, this problem is closely connected with the authorization conflict problem mentioned above. Because such an information disclosure arises when two authorizations having conflict relationship are both allowed.

Papakonstantinou et al. [11] proposed access control authorizations that are used to assign an abstract access control token to RDF triples as specified by means of a query. The method focuses on an RDF triple level, as the syntax process is not sufficient because the remainder of the RDF triples may not, alone, make sense. Knechtel and Stuckenschmidt [12] enforce access restrictions by means of axiom filtering that relies on an axiom labeling. The idea is to label each axiom with a certain access restriction. Users are labeled with the restrictions they are allowed to see. However, to fully support content-based access control, authorization requirements should be set not only on the model's concepts axioms, but also on their individuals. Also, this naive syntactic process is inadequate since the remaining axioms might not make sense alone.

The systems discussed suggest diverse approaches in dealing with the protection of the ontology problem; however, they deal with the protection of the ontology's access concept problem without taking into consideration the authorization decisions according to the content of the data that needs to be accessed, i.e., ABox facts in the domain knowledge. Some of these techniques have assumed the TBox level to be completely public and considered separately outside of the concepts' relationship or concentrated on triples rather than TBox resources. Also, other approaches provided a highly error-prone process with complex mechanisms to secure an ontology knowledge base. Considering the various sensitivities of semantic data in both TBox and ABox paradigms, suitable access control mechanisms pertaining to the semantic repository should be put in place to provide a fully content-based authorization system and combine flexibility as well as a powerful core policy enforcement system, thereby making the process less costly.

Many systems like Jena [13] also attempt to create property like tables out of RDF data. These tables store a number of rows clubbed together on a variety of subjects connected to objects by the same relation. All the RDF triples are distributed into different tables depending on their property. The biggest disadvantage of this mechanism is the sparse representation of RDF data with many NULL values in the property tables which leads to a substantial performance overhead. Another disadvantage with these property tables is the presence of multi-valued attributes in RDF data. For storing multi valued attributes, property tables need lists, sets or other data structures.

V. CONCLUSION

In this paper, the use of authorization control for semantic data repository is described. Our review illustrated at this time is achieved by understanding the basic need of access control policies. All together speaking, the basic need of the authorization control for secure databases and data repositories. In this observation, an overview of semantic data repository is firstly presented for an establishment. For a good authorization control mechanism, the properties of semantic data model are used. From our review, the problem is mainly due to the inference capability of semantic data repositories consequently result in unsatisfactory access control and security attacks.

VI. REFERENCES

[1] Ian Horrocks, Peter F. Patel Schneider, Three Theses of Representation in the Semantic Web, ACM, 2003

[2] Stefan Decker, Frank van Harmelen, Jeen Broekstra,Michael Erdmann, Dieter Fensel, Ian Horrocks , Michel Klein, Sergey Melnik ,The Semantic Web - on the respective Roles of XML and RDF, IEEE 2006

[3] Abdullah Alamri, Peter Bertok, and James A. Thom, Authorization Control for a Semantic Data Repository through an Inference Policy Engine, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 10, NO. 6, NOVEMBER/DECEMBER 2013

[4] The semantic web: Roles of XML and RDF, STEFAN DECKER AND SERGEY MELNIK, FRANK VAN HARMELEN, DIETER FENSEL, AND MICHEL KLEIN JEEN BROEKSTRA MICHAEL ERDMANN IAN HORROCKS,IEEE Internet Computing, October 2000, vol. 15, nr. 3, pgs. 63--74.

[5] Nigel Shadbolt and Wendy Hall, Tim Berners-Lee, The Semantic Web Revisited, Internet Computing, 2006

[6] L. Qin and V. Atluri, “Concept-Level Access Control for the Semantic Web,” Proc. ACM Workshop XML Security, pp. 94-103, 2003.

[7] L. Kagal, T. Finin, and A. Joshi, “A Policy Based Approach to Security for the Semantic Web,” Proc. Int’l Semantic Web Conf., pp. 402-418, 2003.

[8] Abdullah Alamri, Peter Bartok and James A. thom, Authorization control for semantic data repository through an inference policy engine. IEEE transaction on dependable and secure computing, nov 2013, pp 328-341

[9] S. Dietzold and S. Auer, &ldquo,S.: Access Control on RDF Triple Stores from a Semantic Wiki Perspective,&rdquo, Proc. Third European Semantic Web Conf. (ESWC ',06), 2006.

[10] W. Chen and H. Stuckenschmidt, “A Model-Driven Approach to Enable Access Control for Ontologies,” Proc. Int’l Tagung Wirtschaftsinformatik, pp. 663-672, 2010.

[11] V. Papakonstantinou, M. Michou, I. Fundulaki, G. Flouris, and G. Antoniou, “Access Control for RDF Graphs Using Abstract Models,” Proc. 17th ACM Symp. Access Control Models and Technologies, pp. 103-112, 2012.

[12] W. Chen and H. Stuckenschmidt, &ldquo,A Model-Driven Approach to Enable Access Control for Ontologies,&rdquo, Proc. Int',l Tagung Wirtschaftsinformatik, pp. 663-672, 2010.

[13] A. Harth, J. Umbrich, A. Hogan, and S. Decker, “Yars2: A Federated Repository for Querying Graph Structured Data from the Web,” Proc. Sixth Int’l Semantic Web and Second Asian Conf. Asian Semantic Web Conf. (ISWC/ASWC ’07), pp. 211-224, 2007.