 CMC, vol.55, no.3, pp.465-482, 2018

Inverted XML Access Control Model Based on Ontology Semantic Dependency

Meijuan Wang[1, 2, *], Jian Wang[1], Lihong Guo[1, 3] and Lein Harn[4]

Abstract: In the era of big data, the conflict between data mining and data privacy protection is increasing day by day. Traditional information security focuses on protecting the security of attribute values without semantic association. The data privacy of big data is mainly reflected in the effective use of data without exposing the user's sensitive information. Considering the semantic association, reasonable security access for privacy protect is required. Semi-structured and self-descriptive XML (eXtensible Markup Language) has become a common form of data organization for database management in big data environments. Based on the semantic integration nature of XML data, this paper proposes a data access control model for individual users. Through the semantic dependency between data and the integration process from bottom to top, the global visual range of inverted XML structure is realized. Experimental results show that the model effectively protects the privacy and has high access efficiency.

Keywords: Privacy protection, access control, semantic dependence, inverted XML global view.

1 Introduction

In recent years, with the rapid development of big data technology [Feng, Zhang and Li (2014)], database management is more and more important in the whole process of data collection, data analysis, data simplification, accurate and effective data provision, data privacy protection and data release. If all data is hidden or encrypted for privacy reasons, the value of the data cannot be reflected [Meng and Ci (2013)]. Therefore, how to effectively manage and use data safely under the changing and growing needs of big data information structure is particularly important.

1.1 Background

The structure of a traditional database is single, with entity object to be basic unit and the relationship between the entities is clear. Because there is a common model to follow,

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China.

[2] Command & Control Engineering College, Army Engineering University of PLA, Nanjing 210007, China.

[3] Department of Communications Engineering, Nanjing Institute of Technology, Nanjing 211167, China.

[4] University of Missouri, Kansas City, USA.

[*]Corresponding Author: Meijuan Wang. Email: wangmeijuan@nuaa.edu.cn.

there is a model first and then the data. The data management can be top-down approach to achieve integration. However, the current data [Meena (2016)] has a wide range of sources and diverse forms of data representation. Its structure is complex with semantics to be basic unit and the links between entities are unstructured. So, there is the data first and then the model usually. Therefore, it is necessary to follow the bottom-up approach integration with big data fusion [Giuseppe, Josep, Panayiotis et al. (2015)]. The semi-structured and self-descriptive XML language [Wiki (2001)] is a very suitable way to realize this kind of heterogeneous database data exchange and data storage. XML data has also gradually become the common data organization form of database management in big data environment.

Traditional information security focuses on the protection of attribute values [Meng and Ci (2013)]. Through encryption or access control scheme, information of the attribute elements can't be obtained by unauthorized access. As the structure is simple and the rules are simple, the energy consumption of the hardware is small. However, the structural complexity of multi-structured integrated data leads to generally higher cost of implementing encryption schemes [TGakeshi, Blair and Ed (2002)]. Research on XML data security [Wang, Wang, Guo et al. (2018)] had been existed for some time, representative research results still rely on traditional access control methods.

Access control [Michiharu and Satoshi (2000)] is a process for the entity such as user to obtain the necessary permissions to perform the operation. Database of semi-structured data and multi-structured data is a set of nodes, the composition structures of which are graphs rather than tables. So, the traditional operation to achieve the access control program like simple element encryption or selection projection is no longer viable. The original standard language of XML Access is XACML (eXtensible Access Control Markup Language) [XACML Version 3.0], whose advantage is powerful, versatile and open, while the disadvantage is that the architecture is too complex to implement and does not support roles. There are many more advanced authorization mechanisms such as ACLs (access control lists), tiered role based authorization [Moyer and Ahamad (2001)], purpose based authorization [Liu (2015)], rule based authorization [Parmar, Shi and Chen (2002)], and semantic association based authorization [Hulsebosch, Salden, Bargh et al. (2005); Kuang, Deng, Li et al. (2007)], and so on. However, due to the redundant structure of the data itself, the efficiency of these authorizations has not yet reached the desired level. More importantly, the existed research results have been difficult to balance between data value and data security, which means high security reduce the value of the data.

These research results partly do security analysis before operation [Guo, Wang, Wu et al. (2014)] and partly do dynamic judgment [Emami and Zokaei (2010)] during the operational process. However, there are still some unresolved issues:

(1) Fine-grained XML data management lacks flexibility.

(2) Ignoring the semantics of XML elements and associated information of the XML structure, no matter which form of access control programs is likely to result in excessive protection of data and information.

(3) Most existing XML access control theories originate from the improvement of the traditional database model and are always based on the XML structure. In actual operations, the condition judgment caused by the structural redundancy is time-consuming.

One of the reasons that led to the above problems is that it is too dependent on the hierarchical structure of XML itself, and ignoring the direct semantic association nature of data. Therefore, we should combine the semantic nature of big data objects and XML semi-structured data features. The specific content of this paper is the access control scheme based on semantic association semi-structured data, the purpose of which is to efficient and maximum use of public data value for users while meet the needs of the original data privacy protection.

1.2 Motivation

In big data, same-name with different-meanings or different-names with same-meaning is a common form of data aggregation. Before accessing these data, the most important problem to be solved is determining the inline and access basis of the data itself.

On one hand, the formation of big data itself is semantic-based, and then form network by entity links as the main line. So, it is a good way to solve this inconsistency through the semantic relevance. However, most existing XML access control theories originate from the traditional database idea [Parmar, Shi and Chen (2002)], and the structure is the description way, so there is bound to be the loss of semantic meaning, or to maintain the semantics through structural redundancy. What we want to get is to use the simplest structure to describe the most abundant semantic information.

On the other hand, the best way to address the issue of access is empowerment. Common authorization to semi-structured data is structure authorization or node authorization. As for structure authorization, fine-grained data management cannot be done and the information is over-protected easily. As for node authorization, the positioning problem of the node is unavoidable. The existing node positioning is based on the hierarchical path structure and the redundant structure will cause the efficiency bottleneck in condition determination during operation. We hope we can find an authorization way to efficiently obtain data that users can authorize.

Under normal circumstances, data can be authorized to the user will be related to the user's own semantic. The pursuit of data value by users is also based on the existence of semantic association data, not including the data organization. In other words, access control management of XML semi-structured data should focus on the association and authorization of leaf nodes with the largest proportion of information value. Therefore, this paper takes the semantics of the user as the starting point, avoids the original redundancy structure and only associates the data with semantic dependency to the current subject as the valid range accessible by the user. In summary, the main contributions of this paper can be summarized as follows:

(1) Based on the inner semantic relation of data, we mark the user entity by XML Key, and combine the existing XML normalization theory to give the definition and formal description based on entity semantic dependence.

(2) We propose a multi-source data access control model OSACM (Access Control Model by Ontology Semantics) based on user ontology, and propose an algorithm of generating an inverted XML global view based on ontology semantic dependency with the definition of source data privacy. The algorithm can reorganize the accessible objects

that the subject is authorized to obtain effective access rights and reorganize the data views efficiently, so as to improve the access efficiency and provide support for subsequent applications. This is the first time that semantic dependencies have been applied to multi-source access control models.

(3) In order to verify the security and effectivity of this model, we simulate several types of roles of ontology, then analyze its secure View and provide the security verification to ensure the security of the proposed access control precept.

(4) We demonstrate the process of access control and storage of user's View by instances and experiments. The result shows that, the inverted XML global View not only efficiently speeds up the data use but also save the storage space greatly. More importantly, our access control model OSACM addresses the conflict between privacy and data usage value.

1.3 Paper outline

The rest of this paper is organized as follows. Some definitions about semantics and special Key used in our model are described in Section 2. And then we present the overall architecture of the proposed access control model in Section 3. In Section 4, we give the problem analysis and design the most important algorithm for inverted XML Global View based on user's ontology semantics. In Section 5, we show the experiment process and the test result. At last we present the conclusion and summary of this Control Model.

2 Preliminaries

In this section, we formally define some relevant concepts and the example used in this paper.

2.1 Semantic dependency

The KEY in relation database is a very important concept, which decided the integrity constraint. But, the concept of KEY in DTD (Document Type Definition) or XML Schema is not expressive. So, we follow the custom XML Key in Wang et al. [Wang, Bao and Zhao (2007)], and the FD (Function Dependency) and MVD (Multivalued Dependency) in Wang et al. [Wang, Bao and Zhao (2007)] as the basis for further defining the semantic dependencies.

Definition 2.1 (*User Key*) For the XML primary key [Wang, Bao and Zhao (2007)] set K in the DTD, if $e \in K$ and *e* represents the user entity, then *e* is called the user primary key (denoted *uk*), and the set of user primary keys is denoted as UK. Otherwise, we call *e* the entity primary key (remember *ek*), and the entity primary key set is denoted EK.

For XML DTD D and the corresponding XML document $T \models D$, previous research has been carried out based on node value equality FD, and MVD, without considering the semantic dependence. From the perspective of ontology semantics, this paper defines the semantic dependencies based on user primary keys as follows.

Definition 2.2 (Semantic dependency *SD*) Given DTD D and $SD \models D$ δ: Sx~>Sy, δ meets one of the following conditions:

(1) path(Sx)→path(Sy), where Sy function is dependent on Sx;

(2) Sx∈UK, and path(Sy)→→path(Sx). Or, Sy∈UK and path(Sx)→→path(Sy);

(3) If x∈t[path(Sx)], then x∈t[tree(path(Sy))].

Example 1: An example about students selecting electives is given, suppose DTD D1 is as shown in Fig. 1.

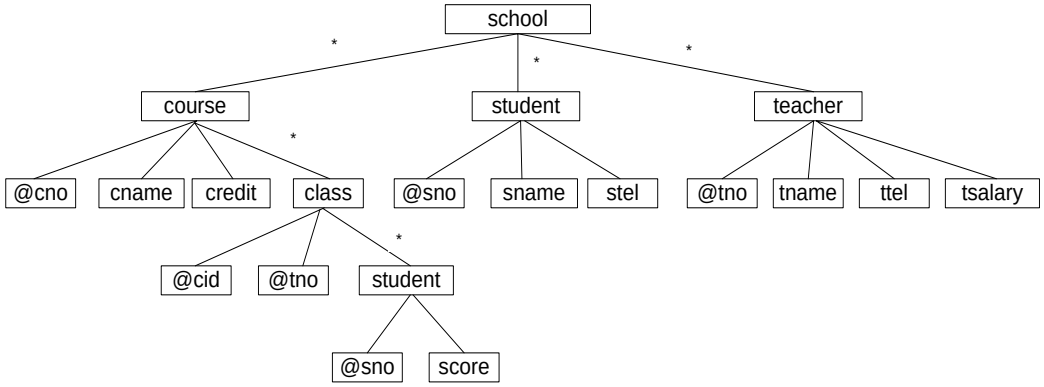

Figure 1: An example D1 about students and electives in school

In D1, the path of a student's score is: *School.course.class.student.score*. When mapped to T1, this path will correspond to multiple paths, representing different values. With respect to D1, the FD, MVD, and the SD can be formalized as:

ϕ1: The Student Number uniquely identifies a student

([school.student.@sno]→school.student)

*ψ*1: The Class ID can identify one or more student

([school.course.class.@cid]→→school.course.class.student)

*δ*1: Course Number semantically depends on Student Number

@sno~>@cno

Different from structural dependence and numerical dependence, the semantic dependence refers to the unstructured dependencies between datas. For example, the a student IDed 'S03' logged in and elected the course 'DB', then δ: @sno['S03']~>@cno['DB'] means that: If the student 'S03' elected course 'DB', then he can gain access to other data which semantically depends on the 'DB' .

2.2 Access control

Definition 2.3 (Access Control Function) The access control function ACC is defined as follow:

ACC: R×Ω×Ā×Ć→α×β (2.1)

The definitions and information of specific notations are clarified in Tab. 1.

Table 1: Notation definitions and information of ACC

Notation	Meaning
R	Sets of all users
Ω	Sets of all data object
Ā/Ū	Sets of user access purposes /Sets of data privacy requirements
Ć	Sets of access control conditions
α	Sets of access codes, α={Yes(Y), No(N)}
β	Sets of operation codes, β={read(r), write(w)}

ACC reflects the restriction of user access intention. That is , for a certain user $r \in R$, under the condition of $\{C_1, C_2,\ldots,C_n\} \in Ć$, and with a certain access purpose $a \in Ā$, ACC defined the privilege of access and operation to every data object $\lambda \in \Omega$. Here the organization of the data itself is invisible to the user, we only emphasis on the representation of the user's access to the data object.

Example 2: For a XML document T1 ⊨D1, the user's basic access control rules can be formalized as follow:

R1: Students can only access information about their own electives and cannot access non-electives

(student,course[*],read,"#login~>@cno")→Y(r), N(w)

R2: Students can only access their own class information and can't access other classes

(student,class[*],read,"#login~>@cid")→Y(r), N(w)

R3: Students can't access other student's information

(student,sudent[*],read,"#login!=@sno")→N(r), N(w)

R4: Students can't access information about teachers' telephone and salary

(student,teacher[ttel,tsalary],read,"NULL")→N(r), N(w)

R5: Teachers can't access other teachers' information

(teacher,teacher[*],read,"#login!=@tno")→N(r), N(w)

Definition 2.4 (Privacy Requirement Function) The data provider's requirement about data privacy is defined in the form of function PRC as follows:

$$PRC: \Omega \times R \times Ū \times Ć \rightarrow \alpha \times \beta \quad (2.2)$$

PRC reflects the data provider's restrictions on data access operations, including the privacy access restrictions of data itself. Compared with ACC, PRC is a dynamically changing permission library. When $u \in Ū$ changes, the algorithm needs to be updated in order to update the user's accessible range.

2.3 Global view

The greatest impact on access control should be the user's "intent". Different roles have different privilege, and different users of the same role have different accessible range.

Definition 2.5 (Semantic Global View) Define the user's semantic global view function UKT as follow:

$$UKT: (ACC+PRC)\times UK\times T\rightarrow Tv \quad (2.3)$$

The definitions and information of specific notations are clarified in Tab. 2.

Table 2: Notation definitions and information of UKT

Notation	Meaning
ACC	Sets of all user's access control function
PRC	Sets of data privacy requirements
UK	Sets of User Key
T	Source data XML document
Tv	The Semantic View of user

UKT function reflects the current user access to the data view. It can be seen that the data objects that can be queried within the scope of the user's access are the semantic related information of the subject and the existing knowledge domain closely related to the individual of the user. The knowledge domain acquisition and data organization will directly affect the data privacy and data usage efficiency. At the same time, we must realize that users are often concerned with the use of data, rather its organization. The huge data structure of the original data access control program is the bottleneck of the implementation, thus making it also the focus of this paper.

3 Access control model by ontology semantics

We proposed the ontology-based access control model shown in Fig. 2. The model is divided into three parts. The first part is source data collection including data structure and also the requirements about source data privacy, which affect the later semantic dependence processing. The second part is a semantic-based access control filter to access different users' view. Its validity, correctness and security will realize the important innovation of this paper. The third part is the mark of the user ontology.

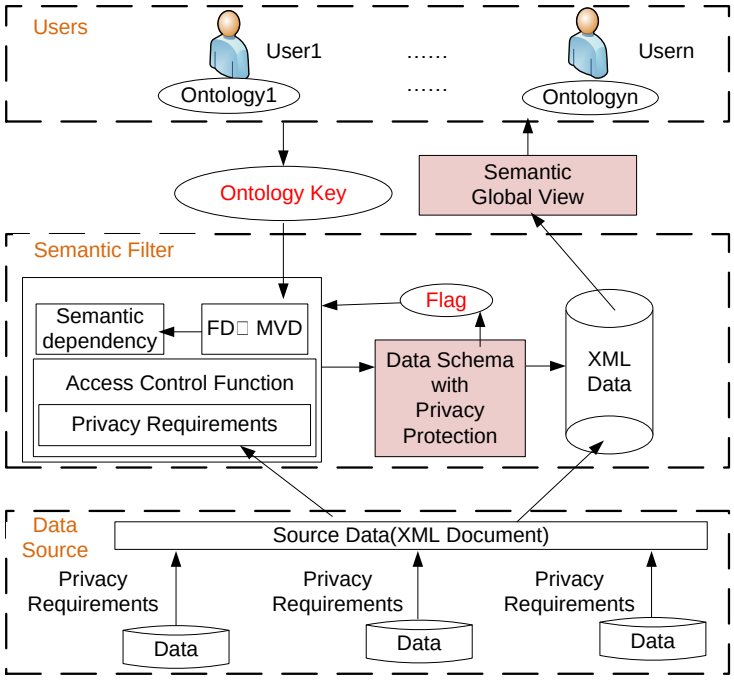

Figure 2: Architecture of access control model by ontology semantics

Access process of OSACM is shown in Fig. 3.

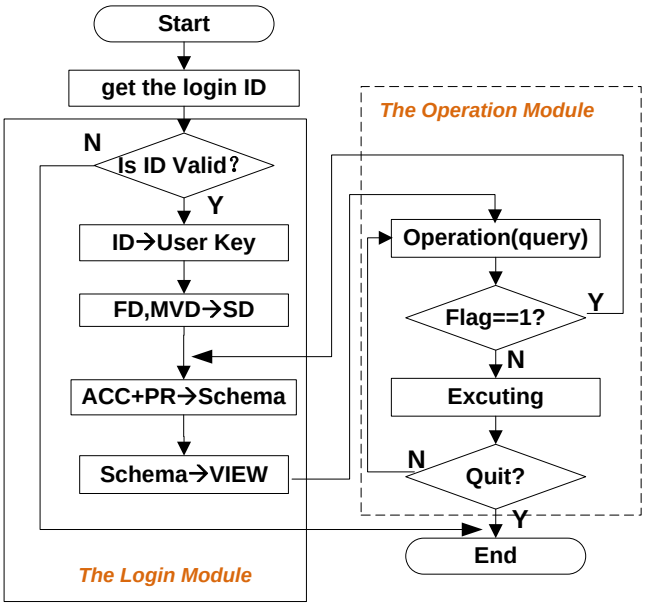

Figure 3: Access process of OSACM

(1) First we mark the user's ontology. Different from previous Role-based access control models, here we simply classify user roles and, more importantly, mark the semantic ontology as **ONTOLOGY**. When obtaining the current ontology User Key, semantic association can be subsequently performed according to the User Key to obtain the global view.

(2) Access Control Filter use the User Key as a start to expand the semantic association, which is a data structure reorganization process. The scheme for generating user-accessible global view through filter is as follows:

Firstly, Semantic Dependency SD of the current User Key is obtained through the FDs, MVDs, and structural dependencies of the XML structure.

Secondly, combined with the formalized access control functions ACC and the user privacy requirement function PRC, the optimized data structure schema is obtained through the algorithm UKSTv, which satisfies the requirement of privacy protection. Here, the most crucial implementation algorithm UKSTv will be discussed in detail in Section 4.

Finally, the global view of the semantic association of the User Key is obtained by matching the source document with the element.

(3) As a multi-source data object, the data contains not only the value of information, but also the data provider's privacy requirements/privacy preferences. When the situation of the data provider changes in different conditions, we can update the privacy requirement effectively according to the Flag mark, so that the global view can be dynamically and efficiently reorganized according to the actual situation of the data source.

4 Inverted XML global VIEW based on semantic dependency

In the process of access control filtering, users first need to acquire their own information knowledge domain (visual range). One of the major innovations in this paper is to generate accessible global views based on user ontology primary keys.

4.1 Problem analysis

According to the principle of "deny priority" [Rota, Short and Rahaman (2010)], the visual DTD global view for students in Example 2 is shown in Fig. 4.

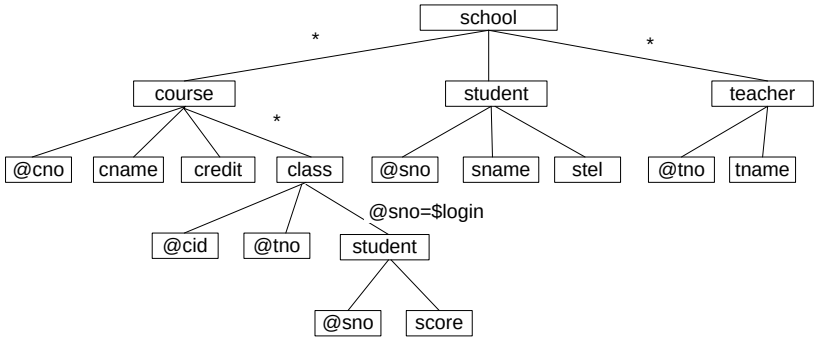

Figure 4: DTD pruning by "deny priority" principle

Because the pruning operation is different from projection, there are three issues:

First: Assuming that the user who is logged in is student number 'S02', according to R1 and R2, the class subtree and course that do not contain "@sno==S02" in the children will be easily judged as inaccessible. But a troublesome problem arises, that the judgment of Xpath "school.course.class.student.@Sno==S02" makes this access control process inefficient.

Second: The semantic orientation of *school.course.class.student* and *school.student* is not clear. When encountering a dependency of "@sno uniquely identifies a student", it is usually discussed from two parts of *school.course.class.student.@sno* and *school.student. @sno*, which is a reasonable redundancy of XML structure data to maintain data consistency. For individual users, such reasonable redundant data not only may be seen as structure redundant, but can also easily lead to confusion in view permissions.

Third: Taking R3 as an example, node 'score' is a valuable data not only constrained by @sno but also by @cno. However, in traditional access control schemes, according to the principle of deny priority, the subtree of *school.course.class.student* is almost pruned, and the 'score' value is overprotected.

Usually users can only see the data associated with himself, some of which are personal attribute data, and other behavioral attribute data. Especially for behavioral attributes, which can be classified as sensitive data, full visibility may cause leakage of reasoning privacy, and the traditional principle of one-size-fits-all may lose the value of data. In fact, it is preferred that 'score' should be under the circumstance that its semantic connection to @sno is broken.

Based on the above analysis, we propose that the global view should be semantically related to ontology, while the semantic dependencies between the data can avoid the structural redundancy and form a new way of organizing. For example, the semantic association can be extracted from the elective information in school as shown in Fig. 5.

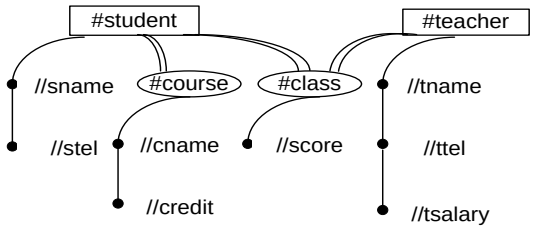

Figure 5: (a) Semantic association about electives by students

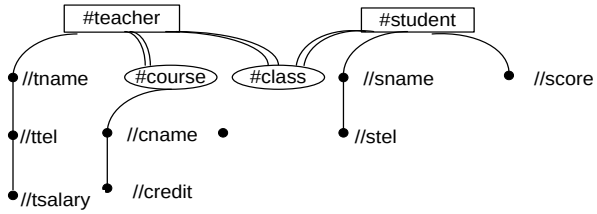

Figure 5: (b) Semantic association about electives by teachers

: Ontology : Object : Include : Associate

From Fig. 5, we demonstrate that our initial analysis and expected goals: Organizational forms of multi-source, unstructured data should respect the integration of modern data semantic associations without over-reliance on redundant hierarchical frameworks.

4.2 Specification

As shown in definition 2.4, if a course number determines a course, when a student gets the ID access of the selected course, he has the access to the data information semantically related to the course. Successively, with the extension of access, we can get a brand new view organization according to access control functions and semantic dependency functions. The semantic association process of the student ontology is shown in Fig. 6.

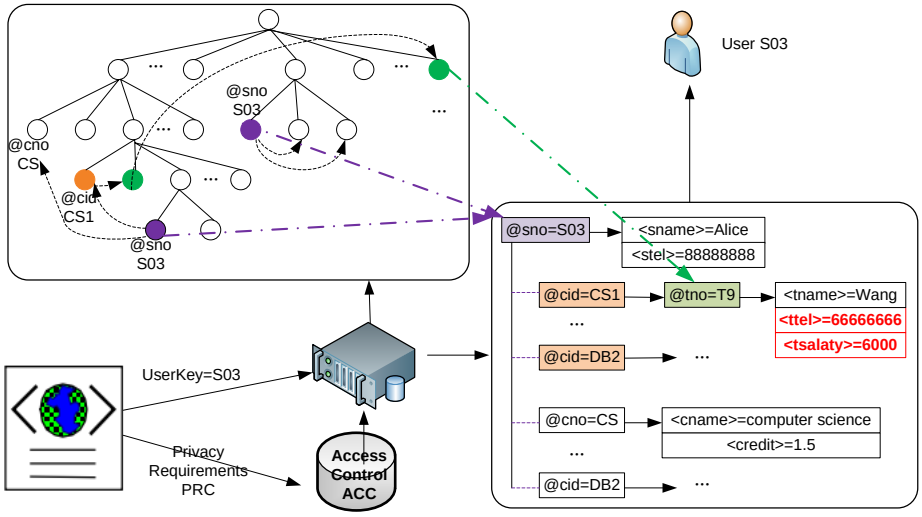

Figure 6: Semantic global view generation process of student ontology

Because the information with the highest weight of the data is in the leaf node, we hope to fully obtain the information of the leaf node without relying on the redundant structural model, extend the authority by the semantic dependency association, and finally obtain the XML view semantically dependent on inverted index. Moreover, considering R4 in Example 2, we can perform further operations as shielding the nodes <ttel> and <tsalary>, which reflects the lightweight flexibility of this paper.

4.3 Specification

From the above analysis and conclusion, this section gives the Algorithm UKSTv. The algorithm UKSTv generate semantic dependency inverted view tree based on the ontology. When the user logs in, the system generates the inverted visual VIEW based on ontology semantics to provide the access of XML data in a more effective and efficient manner, not only meeting the basic requirements of access control but also satisfying the

requirements of privacy protection while providing the maximum data use.

Algorithm UKSTv

Input: T , UK , ACC, PRC, flag

Output: Tv

```
UKStv(T , ACC, PRC, SD , #login)
{
   Tv.root=T.getroot()
  Start：
  //Obtain access control requirement//
    Add PRC to ACC

  //Obtain user attributes associated one-to-one data//
    #login=Tv.element(root , #login , 1)
  for s[i].#login→s[i].Sy in SD
     if ACC(T,#login,read,s[i])==’Y’)
              Sy=Tv.element(root,Sy,1)

  //Obtain user attributes associated one-to-many data//
    for s[i].#login~>s[i].Sy in SD
    if ACC(T,#login,read,s[i])==’Y’)
              Sy=Tv.element(root,Sy,n)

  //Obtain the data associated with the permissions recursively//
```

```
  if(flag==1) //Determine whether there is a dynamic update of source privacy requirements , and return an SD calculation if there is an update
      goto start
  else//Obtain UKT
      node=Tv
      loop:
      if(node==NULL)
        break
      else
         for child in node.child
            for s[i].Sx→s[i].Sy in SD
          if(Sx∈UK)
          if ACC(T,Sx,read,s[i])==’Y’)
             Sy=Tv.element(child,Sy,1)
          else if
                  ACC(T,#login,read,s[i])==’Y’)
                    Sy=Tv.element(child,Sy,n)
      for s[i].Sx~>s[i].Sy in SD
        Sy=Tv.element(child,Sy,n)
     node=child
      goto loop
    return Tv
}
```

Example 3: according to the inference rules in Wang et al. [Wang, Bao and Zhao (2007)], we can get the FD Set and MVD Set of the login student.

Set 1: FD1 Σ’ based on XML Key

school.student.@sno →school.student.sname;

school.student.@sno →school.student.stele;

{school.course.@cno, school.course.class.student.@sno} → school.course.class.student.score;

school.teacher.@tno →school.teacher.tname;
school.teacher.@tno →school.teacher.ttele;
school.teacher.@tno →school.teacher.tsalary;
school.course.@cno →school.course.credit;
school.course.@cno →school.course.cname;
school.course.class.@cid →school.course.class.@tno;

Set 2: MVD1 Π' based on XML Key

school.course.@cno →→school.course.class.@cid
school.course.class.@cid →→
school.course.class.student.@sno

Futher more, according to the User Key *uk*, we can get the SD set of the login student.

Set 3: SD1 Γ' based on login student's User Key

@sno →{sname, stele} @sno~>{@cno,@cid}
@tno →{tname,ttel,tsalary}
@cno →{cname,credit}
{@cno, @sno}→score;
@cid →{@tno} @cid~>{@sno}

At last, according the algorithm UKSTv, we can get the schema tree Tv as shown in Fig. 7 based on the login student ontology.

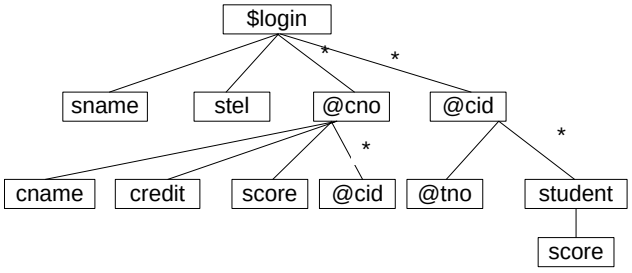

Figure 7: Golbal view schema Tv basesd on the student *uk*

Though there is also a hierarchical structure and a one-to-many subtree relationship in the global View Tv, the schema can be greatly simplified after being mapped to the XML data tree, since only a small number of @cno and @cid are associated with the current login subject semantics as compared to the full set.

Unlike most traditional access control schemes, the algorithm UKSTv retains access available to users to the maximum extent, which improves data availability without over-pruning and affecting data value. For example, the nodes of original Xpath

school.course.class.student.score can still be used. For the sensitive data *score*, deleting the student subtree in order to avoid the semantic association between the score and the subtree @sno will result in excessive data protection, so we want to maximize the original inline value of the data.

5 Experiment result and analysis

5.1 Experiment preparing

The experiments in this paper were performed on the I7 processor 2.81 HZ, memory 8 G, 64-bit Windows 10 operating system, by XML Spy and Python 3.6 programming tools. The analysis of the XML document of the test set [XML Data] in the experiment adopts the DOM technology to acquire the node information and content information. The test data set and related attributes are shown in Tab. 3.

Table 3: Test dataset information

Number	Dataset name	Size, KB	Nodes	Depth
1	own.xml (self-defined)	5	138	5
2	yahoo.xml	24	342	5
3	reed.xml	283	10546	4
4	uwm.xml	2338	74557	4

This paper compares with the classical access control model ViewAC is proposed by Damiani et al. [Damiani, Vimercati, Paraboschi et al. (2002)]. ViewAC is also a view-based access control mechanism that implements privacy protection in combination with access control methods. It pre-defines and saves stand-alone access views based on a class of roles. In the process of view creation, privacy nodes containing sensitive information are tailored. Our OSACM scheme is based on the user ontology, not only consider the access control requirements, but also consider the privacy needs of the data provider. We can obtain a semantic global view of the inverted index that simplifies the structural redundancy. There are plans for comparison of storage and query time with original algorithm in different scenarios.

5.2 Experiment preparing

In order to test the storage space and execution efficiency of this project, we mainly test from the following aspects: (1) different users' storage space under different test sets; (2) access (query) response time of different users under different test sets. In the ViewAC scheme, XPath positioning is chosen as the XML structured query language. In the scheme of this paper, the semantic dependence function is used to solve the positioning.

Fig. 8 respectively shows the space ratios of the four test sets in one user view and three user views. The abscissa represents the data set number, and the ordinate represents the storage space in KB.

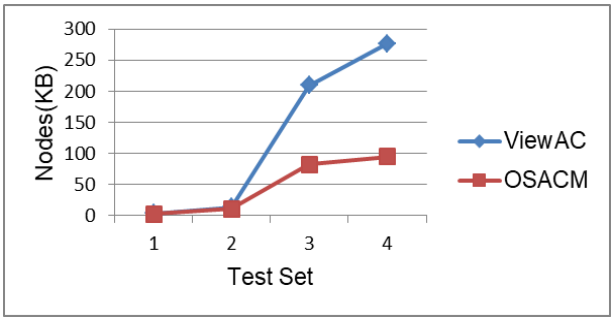

Figure 8: (a) Comparison of storage space in one view

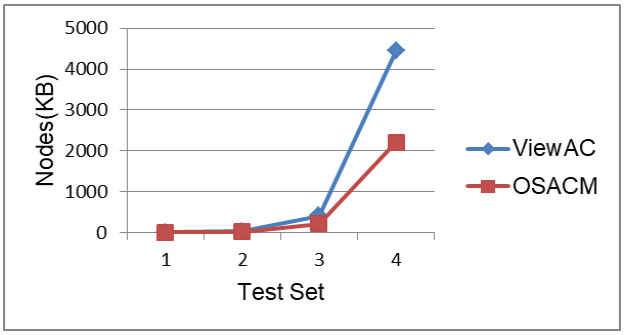

Figure 8: (b) Comparison of storage space in three views

Fig. 9 shows the comparison of the security view loading time for the four test sets in Fig. 8(a) when the user logs in, and the ordinate represents the response time in seconds. It can be seen that the performance of the loading time of the security view is not directly proportional to the scale of the original data. It is also affected by the constraints of the scale of the rule conditions and the complexity of the structural model. What happens in Test Set 4 is that valuable information cannot be accessed with fine-grained because of the overall pruning. Although it saves response time in the case of guaranteeing security, it affects data availability.

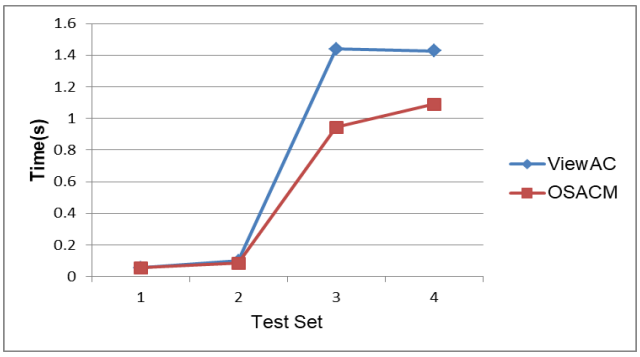

Figure 9: Comparison of resolution time by user of Fig. 8(a)

Some query time results with different view of different ontology are shown in Tab. 4.

Table 4: Query responsed time of test XML document

Number	Dataset name	Query	Search time, s
1	own.xml	/root/course/class/student/score	0.009322
		$score[@cno=’C02’]	0.008454
2	own.xml	avg(root//score[@cno=’C02’])	Null
		ave(score) [@cno=’C02’]	0.029287
3	yahoo.xml	T	1.482326
		Tprint(T)	1.324688
4	yahoo.xml	/root//listing/auction_info/current_bid[>50]	0.068274
		//current_bid[>50]	0.052502
5	reed.xml	T	6.233467
		Tprint(T)	5.438899
6	uwm.xml	/root/coutse_listing[course=’216-088’]/section_listing//section	0.975565
		section[course=’216-088’]	0.654328
7	uwm.xml	/root/coutse_listing[course=’360-888’]/section_listing//section	2.089764
		section[course=’360-888’]	1.871186

From the test data and experimental results, we can know that when the direct semantic relationship between parallel data is clearly defined, the larger the complex structure is, the higher the efficiency of the implementation of the project is. Meanwhile, the deeper the structure, the higher the efficiency of implementation. Because high-valued leaf data is directly obtained by semantic functions and no longer depends on lengthy path judgment. While in the case that the user's access does not reflect the individual’s situation in general, the efficiency of the implementation of this project is not obvious, but it has obvious advantages in improving the value of information to prevent excessive protection of information.

In addition, combined with the structure of the experimental data set, it can be analyzed that the number of leaf nodes and the size of the access-based privacy constraints rule have a great influence on the model, while the structure characteristics of the original XML document have little impact on the query efficiency under our security view.

6 Conclusions

Nowadays, data diversification and explosive increment have brought new demands and challenges to database management. Theories and technologies on data integration, data analysis and data mining are continuously updated, making the privacy protection of information once again the focus of users' attention. There is a model first and then the

data in the traditional database, operation management can be top-down, property positioning and security operations are easy to achieve. And big data are often semantics-based collection of entities semantic relations between the staggered into a network, usually there is the data first and then the model, operational management needs big data fusion technology for bottom-up, so, it is difficult to have a unified and standard data management.

XML database as the big data carrier has semi-structured, self-descriptive and other advantages. Although there exist some research results based on XML secure access, they are still based on the user defined Schema structure. The fine-grained operation depends on the positioning of XPath, which not only costs large for implementation but also ignores the original semantic relationship of data.

Considering that the user's acquisition of data value is not concerned with the form of data organization, the user's authorization should combine the semantics of the data itself and its association. In this paper, in view of the contradiction between data value and data privacy protection in the era of big data, we propose an access control model based on user ontology and a semantic-dependent inverted global view XML algorithm (UKSTv), which avoids redundant structure association and keeps flexible and controllable semantic associations. The data of the simulation experiment shows the good performance of the scheme and can provide the basis for the efficient and safe data usage.

Future work will include further research on how to implement more flexible update and maintenance on the inverted global view of XML, and the prevention of similar inference attacks based on the dynamic context of the network.

Acknowledgement: This work was supported by Funding of Jiangsu Innovation Program for Graduate Education KYLX_0285, the National Natural Science Foundation of China (No. 61602241), the Natural Science Foundation of Jiangsu Province (No. BK20150758), and the pre-study fund of PLA University of Science and Technology.

References

Damiani, E.; Vimercati, S. D. C. D.; Paraboschi, S.; Samarati, P. (2002): A Fine-grained access control system for XML documents. *ACM Transactions on Information Systems and Security*, vol. 5, no. 2, pp. 169-202.

Emami, S. S.; Zokaei, S. (2010): Context-sensitive dynamic role-based access control model for pervasive computing environments. *ISC International Journal of Information Security*, vol. 2, no. 1, pp. 47-66.

Feng, D. G.; Zhang, M.; Li, H. (2014): Big data security and privacy protection. *Chinese Journal of Computers*, vol. 37, no. 1, pp. 246-258.

Giuseppe, D. A.; Josep, D. F.; Panayiotis, K; Vicen, T.; Yves-Alexandre, M. et al. (2015): Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. *European Union Agency for Network and Information Security (ENISA) Report*.

Guo, L. H.; Wang, J.; Wu, H. T.; Du, H. (2014): eXtensible Markup Language access control model with filtering privacy based on matrix storage. *Institution of Engineering*

and Technology, vol. 8, no. 11, pp. 1919-1927.

Hulsebosch, R. J.; Salden, A. H.; Bargh, M. S.; Ebben, P. W. G.; Reitsma, J. (2005): Context sensitive access control. *Proceedings of the SACMAT*, vol. 11, no. 41, pp. 111-118.

Kuang, L.; Deng, S. G.; Li, Y.; Wu, J.; Wu, Z. H. (2007): Using inverted indexing to facilitate composition-oriented semantic service discovery. *Journal of Software*, vol. 18, no. 8, pp. 1911-1921.

Liu, Y. M. (2015): *Research on Privacy Data Access Control Mechanism Based on Access Purpose (Ph.D. Thesis).* Fudan University.

Masayoshi, T.; Issei, Y.; Naohiko, U. (2008): Rule-based XML mediation for data validation and privacy anonymization. *Proceedings of the IEEE International Conference on Services Computing*, pp. 21-28.

Meena, S. D. (2016): Data lakes-A new data repository for big data analytics workloads. *International Journal of Advanced Research in Computer Science*, vol. 7, no. 5, pp. 65-66.

Meng, X. F.; Ci, X. (2013): Big data management: Concepts, techniques and challenges. *Journal of Computer Research and Development*, vol. 50, no. 1, pp. 146-169.

Michiharu, K.; Satoshi, H. (2000): XML document security based on provisional authorization. *Proceedings of the ACM Conference on Computer & Communications Security*, vol. 279, no. 7, pp. 87-96.

Moyer, M. J.; Ahamad, M. (2001): Generalized role-based access control. *Proceedings of the International Conference on Distributed Copmputing Systems*, pp. 391-398.

Parmar, V.; Shi, H. C.; Chen, S. S. (2003): XML access control for semantically related XML document. *Hawaii International Conference on System Sciences*, vol. 9, no. 9, pp. 9-10.

Rota, A.; Short, S.; Rahaman, M. A. (2010): XML secure views using semantic access control. *Proceedings of International Conference on Extending Database Technolog*, pp. 1-10.

TGakeshi, I.; Blair, D.; Ed, S. (2002): XML encryption syntax and processing. http://w3.org/TR/2002/CR-xmlenccore-20020802.

Wang, M. J. (2007): *Normalized Storage of XML Data (Ph.D. Thesis).* Nanjing Normal University.

Wang, M. J.; Bao, P. M.; Zhao, G. L. (2007): Translating XML to normal relational schema based on XML key. *Computer Science*, vol. 34, no. 3, pp. 95-97.

Wang, M. J.; Wang, J.; Guo, K. J. (2018): eXtensible markup language keywords search based on security access control. *International Journal of Grid and Utility Computing*, vol. 9, no. 1, pp. 43-50.

Wiki (2001): XML. https://en.wikipedia.org/wiki/XML.

XACML Version 3.0. https://www.oasis-open.org/committees/xacml/.

XMLData Repository. http://www.cs.washington.edu/research/xmldatasets/. University of Washington XML Repository.