IPSJ Transactions on Bioinformatics Vol. 1 42–52 (Nov. 2008)

Original Paper

A Grid-aware Access Control Mechanism in a Clinical Database for Parkinson's Disease Research and Diagnosis

Susumu Date,[†1] Kazunori Nozaki,[†1]
Haruki Nakamura,[†2] Saburo Sakoda[†3]
and Shinji Shimojo[†1,†4]

The Grid has increasingly gathered the attention and interest of scientists and researchers as a building block technology for computational infrastructure. Because of the recent increasing demands on the Grid, the utilization of the Grid has been explored in various scientific research areas. In reality, however, the Grid is not fully utilized well in today's practical scientific research areas treating security-sensitive data, especially in biomedical research areas. This is partly due to the lack of know-how about how access control can be achieved in the actual applications despite the maturity of security technologies related to authentication and authorization. From this perspective, in this paper, we present a Grid-aware access control mechanism leveraging MyProxy, GSI, PERMIS and XSLT, which we have built into a clinical database for Parkinson's research and diagnosis. In particular, we focus on how these technologies are used to satisfy the access control requirements derived from a clinical database. Also, it is shown that the proposed access control mechanism can be operated with low-cost administration and acceptable overhead.

1. Introduction

Recently, the Grid has increasingly gathered the attention and interest of scientists and researchers as a building block technology for computational infrastructure. Because of increasing demands on the Grid, the utilization of the Grid has been explored in various scientific research areas. Examples of such scientific research areas include high-energy physics, bio-science, material science, medicine and even social science. Due to much effort by Grid researchers so far, the Grid is becoming a convenient and easy-to-use technology more and more for scientific research. In reality, however, the Grid is not utilized well in today's actual scientific research areas because of security problems.

Generally, in a Grid environment, many users with various user attributes are supposed to utilize a diversity of computational and data resources. For this reason, an access control solution that forces users to access such resources properly depending on the users' attributes is essential. Moreover, an access control system becomes important especially in scientific research areas, such as biology and medicine, which treat security-sensitive data.

In general, access control is performed through the two operations of authentication and authorization. Authentication is an operation of verifying the identity of the individuals, and authorization is an operation that determines what kind of actions can be permitted to the identified individuals. For authentication, the Globus grid toolkit [1], a de facto standard implementation of the Grid, has provided Grid Security Infrastructure (GSI) [2]. Today, GSI is widely accepted and utilized. On the other hand, for authorization, many technologies such as CAS [3], VOMS [4], Akenti [5] and PERMIS [6] have been proposed and implemented. However, how these technologies can be used for the access control function in actual applications, and how these technologies can effectively work in the access control function are still being explored. Accordingly, the know-how and techniques about how authentication and authorization technologies are utilized in realizing an access control function that satisfies the requirements from actual applications should be discussed for the use and promotion of the Grid in scientific research.

In this paper, we detail the architecture of the access control mechanism built into the clinical database for Parkinson's disease research and diagnosis briefly overviewed in Ref. 7). Also, we evaluate and discuss the architecture from aspects of practical use and administration.

The rest of this paper is organized as follows. Section 2 presents an overview of the clinical database which we have been developing. In Section 3, we consider the requirements for access control to the database. Next, in Section 4, we detail the access control mechanism built into the database and then, in Section 5, we evaluate the usefulness of the access control mechanism. In Section 6, we review

†1 Cybermedia Center, Osaka University
†2 Institute for Protein Research, Osaka University
†3 Graduate School of Medicine, Osaka University
†4 National Institute of Information and Communications Technology

related works. We conclude this paper in Section 7 with suggestions for further research and development.

2. Clinical Database for Parkinson's Disease Research and Diagnosis

2.1 Parkinson's Disease and the Necessity of Clinical Database Research and Diagnosis

Parkinson's disease is a degenerative brain disorder. This disease is characterized by muscle rigidity, tremors, bradykinesia (a slowing of physical movement), and in extreme cases, a loss of physical movement. Also, this disease occurs when certain neurons in a part of the brain called the substantia nigra die or become impaired [8]. An aging society with age-related diseases such as Parkinson's is emerging in most developed countries.

However, effective treatments and remedies against this disease have not been established yet to date although the cause of this disease is considered to be the insufficient formation and action of dopamine, which plays a role of great importance in brain activity. Obviously, a major reason for this current situation is explained by the fact that the functional mechanisms of human brain are still unknown despite neuroscientists' efforts.

Furthermore, research and clinics related to Parkinson's disease have not been fully supported by information technology. In reality, therapy and remedy relying on dopamine-based drug dosing have been performed in most clinics and hospitals. Nonetheless, the clinical data and the knowledge and know-how about the therapy and remedy, such as what kind of dopamine-based drug was effective for what type of patients, have not been shared and used among these clinics and hospitals despite the recent advancement and maturity of networking and computer technology. Therefore, efficient statistical analysis of clinical data cannot be performed today although it is considered to be useful and effective towards an understanding of Parkinson's disease and the establishment of its treatment and remedy.

2.2 Concept of the Clinical Database

Due to the increasing demand for a clinical database for Parkinson's disease research and diagnosis, we have been developing such a database in hope that the clinical database is securely and seamlessly shared and used in a cross-organizational manner. **Figure 1** shows the concept of the clinical database. The user model behind the concept was simplified for this preliminary research. This concept assumes that medical doctors, co-medicals and their research collaborators utilize a clinical database for Parkinson's disease. Also, medical doctors and co-medicals, such as nurses and X-ray technicians, are assumed to work at a hospital where the clinical database is managed, while the collaborators are assumed to work at different organizations such as clinics and research laboratories.

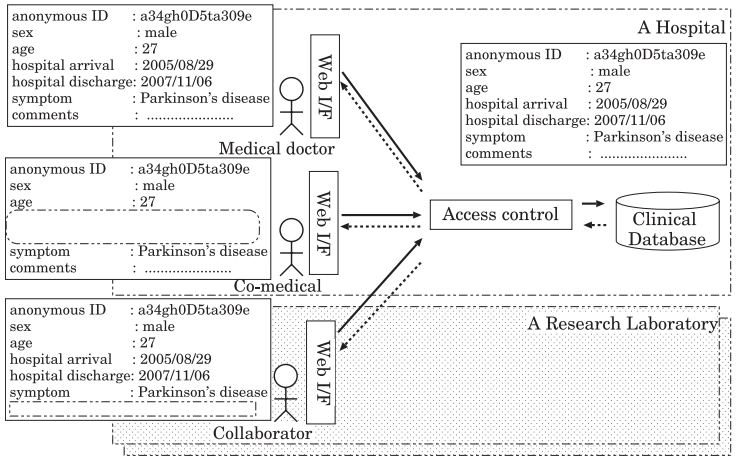

Fig. 1 Overview of clinical database.

The idea of the clinical database is very simple in that it simply and securely provides the user with appropriate clinical data depending on the user's attributes. Therefore, clinical data provided by the clinical database looks different to the user depending on whether or not the user is a medical doctor, co-medial or collaborator.

2.3 Data Format for Description and Sharing of Clinical Data

To realize the sharing of clinical data on Parkinson's disease among multiple clinics and hospitals, a standard data format for describing Parkinson's disease symptom and patient information is indispensable. However, such a standard format does not exist in the current clinical environment. The reason for this lack

of a standard format is because no effective therapies and remedies for Parkinson's disease have been found and established yet as described in Subsection 2.1. In fact, in Osaka University Hospital, a proprietary data format was structured and clinical data is managed as a file conforming to the format.

We have newly defined a MML (Medical Markup Language)-based data format for describing and sharing clinical data pertaining to Parkinson's disease in this preliminary research, in the hope that our study is a candidate for a standard format. MML [9] is a XML-based data format which conforms to Health Level 7 (HL7), a standard for clinical information exchange [10]. **Figure 2** shows an actual part of clinical data format conforming to MML. In this data format, each piece

```
<?xml version="1.0" encoding="UTF-8"?>
<levelone xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:noNamespaceSchemaLocation="H:\parkinson\baseData.xsd">
             .....
             .....
<personal>
   <patient_info>
      <initial>
         <firstName_init>H</firstName_init>
         <familyName_init>*</familyName_init>
      </initial>
      <birthday>
        <birthday_yy></birthday_yy>
        <birthday_mm></birthday_mm>
      </birthday>
      <sex>female</sex>
   </patient_info>
   <doctor_info>
     <doctor_type code="DT001">
   </doctor_info>
   <update_userId>UU002</update_userId>
   <personal_id>03</personal_id>
</personal>
             ....
             ....
<medicalHistory_part history_code="199000">
  <medical_info>
    <medical_type>MT001</medical_type>
    <medical_date>
       <medical_date_yy>1999</medical_date_yy>
       <medical_date_mm>11</medical_date_mm>
       <medical_date_dd>**</medical_date_dd>
    </medical_date>
    <medical_detail>
      The patient was aware of having difficulty in extending the
      patient's left maniphalanx around 1999 and had massaged since
      then. There was not improvement in symptoms.
    </medical_detail>
    <hospital_name>
      Osaka University Hospital
    </hospital_name>
    <hospital_section>
      Department of Neurology
    </hospital_section>
    <firstExam_date>
             ....
             ....
```

Fig. 2 Patient clinical data in MML.

of the clinical data is composed of the patient's information such as name and age, the results of diagnosis, dosing information, and hospital admission and discharge dates, described as an XML element. These elements have different data security attributes. For example, medical doctors are permitted to see all elements while their research collaborators at other organizations are permitted to see only part of the dosing information.

3. Requirements to Access Control

In this section, we clarify three requirements toward the access control of the clinical data, which we have determined through discussions from both the fields of computer science and medical research and from clinical points of view.

3.1 Interoperability

The amount of medical data to analyze has been dramatically increasing recently because of the advancement of measurement technology and computational technology. In fact, medical measurement devices such as MEG (Magnetoencephalography) and fMRI (functional Magnetic Resonance Imaging), has improved the accuracy of measurement in both temporal and spatial resolution year by year. Also, high-performance computing technology has been dramatically advancing. Due to the technological advancement, with cluster computing and Grid technology, scientists can now combine the computational power brought together by thousands of computers even if they are geographically distributed [11].

Greater computational power is used in many fields of science today, but its use is still rather remarkable in medical science. Today, medical analysis methods, such as signal processing of brain functional data and volume rendering of the bronchial tubes, requires more computational power than can be achieved through a single computer. For this reason, high-performance computing technologies are being implemented more and more in many analysis methods. For example, medical analysis and related life science software are about to be implemented as Web services and Grid services available through the Internet [12)–14)].

Our clinical database, therefore, should have an interoperation functionality with such medical analysis software in the near future. An example of such interoperation is the direct transfer of result data of such medical analysis to the

database. For this reason, the clinical database we envision requires interoperability with HPC-aware medical software and services. Moreover, the access control mechanism for the clinical database has to be able to seamlessly interoperate with such HPC-aware software in terms of authentication and authorization.

3.2 Fine-grained Access Control and Data Filtering

In a cross-organizational environment where multiple organizations such as clinics, hospitals, and research laboratories are coupled on the Internet, various users such as researchers and medical doctors are supposed to access data of their interest. Furthermore, the data itself have their own data security requirements based on their confidentiality level. The following example explains this situation in detail.

In the case shown in Fig. 1, medical doctors and co-medicals working for hospitals, and researchers who work for Parkinson's disease research are assumed to want to access our database for their research and clinical purpose. In this case, medical doctors have to fully access the clinical data pertaining to their patient stored in the database located at the hospital in which they work, while medical doctors who do not diagnose the patient are not allowed to access the security-sensitive part of clinical data. Such security-sensitive data include patients' privacy information such as name, age and address. Furthermore, co-medicals, such as nurses and X-ray technologists, who work for the medical doctors in charge of the patient at the same hospital are not allowed to access the medical doctors' clinical or research comments and notes about patients and patients' symptom, while researchers who work for other hospitals and collaborate with the medical doctor are allowed to access such information and note for statistical research.

To realize such a clinical database that can be shared in a cross-organizational manner, a fine-grained access control that judges whether or not to permit access to clinical data not on a file basis but on a data element level depending on user's attributes is essential. Moreover, a fine-grained data filtering mechanism that offers a set of data elements appropriate to the user who accesses the database is important in building such a database.

3.3 Low-cost Administration of Security Policy and Configuration

To keep the database secure, administering its security policy and configuration on which access control judgment is performed so that it is always up-to-date to the change of user attribute is of great importance. For example, in the case that a medical doctor transfers from one hospital to a different hospital, the administrator of the database may have to immediately reflect the change in user attribute of the medical doctor to the system in response to the event. Also in the case that a researcher in a research laboratory stops collaborative research with a medical doctor working for a hospital for some reason, the administrator promptly has to change the access control configuration to prevent the researcher from accessing the data stored in the database located at the hospital. This kind of change in user attributes occurs frequently in the current clinical and research situation. Thus, low-cost administration of a security policy and configuration is an important factor in building the access control mechanism for the database from an administrative point of view.

4. Access Control Mechanism

In this section, the access control mechanism which we have built for the clinical database is explained.

4.1 Access Control Model

In this preliminary research, a simple access control requirement model for the database has been set. **Table 1** summarizes the simple access requirement model. "o" and "x" means "accessible" and "not accessible". In this model, the database users are categorized into the three classes of medical workers: doctors (Doctor) who work for the Department of Neurology at Osaka University, co-medicals (InnerUser) who work for the medical doctors, and research collaborators (OuterUser) who work at other hospitals and research laboratories. Each user class is allowed to access a specific set of data elements written in MML, based on the model shown in Table 1. Under this model, medical doctors and co-medicals have the privilege of accessing data elements related to patient privacy,

Table 1 Access control model for the clinical database.

	Privacy Info.	Patient-inferred Info.	Clinical Info.
Doctor	o	o	o
InnerUser	o	o	o
OuterUser	x	x	o

data elements from which patients can be inferred, and the data elements related to medical doctors' diagnosis, while research collaborators do not have the privilege to access data elements related to patient privacy and the data elements from which patients can be inferred.

4.2 Access Control

Figure 3 shows the access control mechanism we have developed into the clinical database. In this research, we have adopted Grid-aware technologies based on the interoperability requirement with HPC-aware medical analysis software to develop the clinical database system. For this reason, GSI (Grid Security infrastructure), a technology that provides a single sign-on authentication service based on PKI (Public Key Infrastructure) using X.509 public key certificate for user authentication [2),15),16)] and PERMIS (Privilege and Role Management Infrastructure Standard), which has a high affinity with Grid technologies, have been used for authentication and authorization on top of the Globus toolkit 4.0.1.

GSI allows users to access a variety of resources on the Grid without forcing them to show their credentials such as passwords and passphrases many times. On the other hand, PERMIS is a technology that provides an authorization method for allowing/denying users' access to computational and data resources

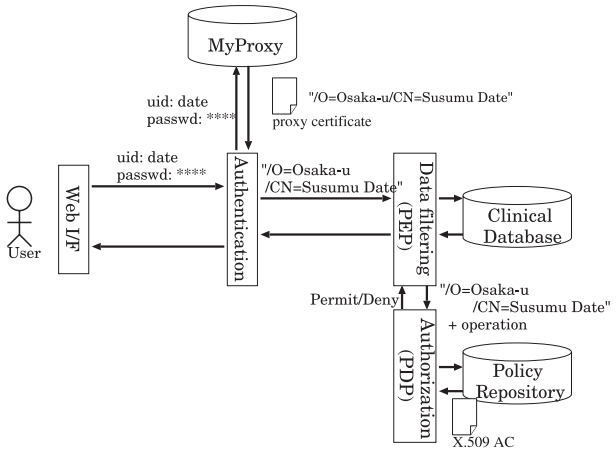

Fig. 3 Access control mechanism.

based on PMI (Privilege Management Infrastructure) [6),17)]. PMI is an authorization infrastructure concept similar to the concept of PKI as an authentication infrastructure. Under PERMIS, several roles such as secretary, manager, and employee are defined; a set of permissions is assigned to each role, and each user is assigned to one or more roles. Based on this role assigned to the user, PERMIS determines whether the user is permitted to invoke an action, or a Java method of the Grid service developed on top of Globus grid toolkit. In PERMIS implementation, X.509 attribute certificate is used to describe the user's role and the permissions given to the user.

By combining these two technologies, our access control mechanism performs the judgment of whether the user can invoke a method of Java program implemented on the Globus toolkit or not. The access control mechanism works as follows. First assume that a user whose global name (distinguished name used in the certificate) is "/O=Osaka-u/CN=Susumu Date" accesses the clinical database. At this time, our access control mechanism first attempts to verify the identify of the user using GSI. In our system, MyProxy [18)] which is an online credential system that manages user's credential on behalf of the user has been adopted. In this process of GSI authentication, the MyProxy receives a pair of user IDs and passwords and then feeds the user credentials registered in advance (the user's proxy certificate [19)]) which is used for subsequent authentication processes on the Grid. After that, the GSI authentication module accesses the Data filtering module as the user "/O=Osaka-u/CN=Susumu Date" by using the user proxy certificate. Then, after completing mutual authentication processes with the authentication module, the Data filtering module asks whether or not to permit access to the clinical database on behalf of the user. In turn, PERMIS authorization module as a Policy Decision Point determines whether to permit the user's access by retrieving the attribute certificate corresponding to the user role from the attribute certificate repository immediately after receiving the query. Finally, the Data filtering module as a Policy Enforcement Point accesses the clinical database on behalf of the user if permitted.

Figure 4 shows an example of the X.509 attribute certificate stored in the attribute certificate repository. This certificate indicates that the user with a Doctor role can invoke the getDoctorXSLT() method running at

```
<X.509_PMI_RBAC_Policy OID="2006.08.19.20.20.01">
  <SubjectPolicy>
    <SubjectDomainSpec ID="Neurology">
      <Include LDAPDN="OU=neurology,O=med.osaka-u,C=jp"/>
    </SubjectDomainSpec>
  </SubjectPolicy>

  <RoleHierarchyPolicy>
    <RoleSpec OID="1.2.826.0.1.3344810.1.1.14" Type="permisRole">
      <SupRole Value="Doctor"/>
      <SupRole Value="OuterUser"/>
    </RoleSpec>
  </RoleHierarchyPolicy>

  <SOAPolicy>
    <SOASpec ID="TheSOA" LDAPDN="CN=Permis Soa,O=biogrid,C=jp"/>
  </SOAPolicy>

  <RoleAssignmentPolicy>
    <RoleAssignment>
      <SubjectDomain ID="Neurology"/>
      <RoleList>
        <Role Type="permisRole" Value="Doctor"/>
      </RoleList>
      <Delegate Depth="0"/>
      <SOA ID="TheSOA"/>
      <Validity/>
    </RoleAssignment>
  </RoleAssignmentPolicy>

  <TargetPolicy>
    <TargetDomainSpec ID="DataFilteringService">
      <Include URL="https://133.1.33.***:8443/wsrf/DataFilteringService"/>
      </TargetDomainSpec>
  </TargetPolicy>

  <ActionPolicy>
    <Action Args="" Name="getDoctorXSLT"/>
    <Action Args="" Name="getInnerXSLT"/>
    <Action Args="" Name="getOuterXSLT"/>
  </ActionPolicy>

  <TargetAccessPolicy>
    <TargetAccess>
      <RoleList>
        <Role Type="permisRole" Value="Doctor"/>
      </RoleList>
      <TargetList>
        <Target Actions="getDoctorXSLT">
          <TargetDomain ID="DataFilteringService"/>
        </Target>
      </TargetList>
    </TargetAccess>
  </TargetAccessPolicy>
</X.509_PMI_RBAC_Policy>
```

Fig. 4 Attribute certificate.

http://133.1.33.****/wsrf/DataFilteringService to access the clinical database.

4.3 Fine-grained Data Filtering Module

By utilizing the functionality of the PERMIS Java method level invocation, we have built a data filtering function that interoperates with the access control mechanism so that our database provides fine-grained access to MML-based clinical data for Parkinson's disease research. **Figure 5** shows the data filtering

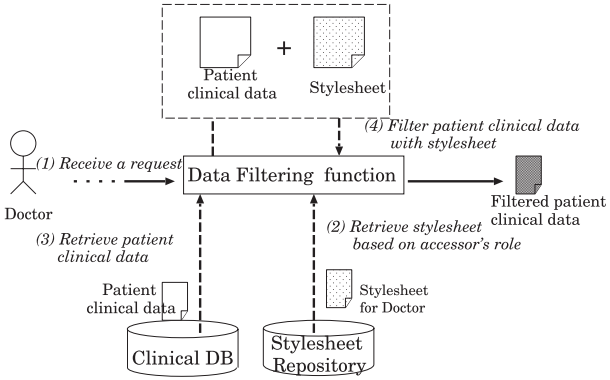

Fig. 5 Data filtering module.

module in action. The core feature of our data filtering module is characterized by the use of XSLT (eXtensible Stylesheet Language Transformation). Originally, XSLT is a protocol that converts a XML document to other XML document based on a stylesheet written in XSL and commonly used in e-commerce such as enterprise data exchanges. The data filtering function applies an XSL file to the original MML-based clinical data based on the user's role to extract a set of MML data elements accessible to the user. For this purpose, either the Java method of getDoctorXSLT(), getInnerXSLT(), or getOuterXSLT() is implemented based on Doctor, InnerUser, or OuterUser role, respectively, through the function of PERMIS. Each of these methods can be used by the corresponding user role. As a result, data elements related to patient privacy and data elements from which a patient is inferred, such as hospital admission and discharge dates, are removed for access from research collaborators at other hospitals and research laboratories.

5. Evaluation and Discussion

In this section, we evaluate and discuss the usefulness of the access control mechanism. For this purpose, we first measure the overhead of the access control mechanism and evaluate how the measured overhead has an impact on the actual use of the clinical database. Secondly, the administration cost of the access

control is discussed.

5.1 Overhead Measurement

For the measurement, the following simple experimental environment was set up. Although the authentication module, the access control module, and the data filtering module can each be deployed on separate computers, these components were deployed on a computer running Linux Kernel 2.6 with an Intel Pentium 4 2.8 GHz processor and 1 GB memory. Also, a computer running Microsoft Windows 2000 with an Intel Pentium 4 2.8 GHz and 1 GB memory was used to accommodate actual patient clinical data.

In this experiment, the time taken for the access control mechanism to receive a request from the user and then obtain a stylesheet appropriate for the user's role (process (a)), and the time taken for the access control mechanism to search the clinical database for a MML-based clinical data and then complete the data filtering process (process (b)) were measured. For measurement, we accessed the clinical database as either a user with the Doctor role or a user with the OuterUser role. To remove the search time for clinical data of interest at the XML database, we have located only a few clinical data in an identical directory on the Windows machine.

Table 2 shows the results of the measurement. The results indicate that the time taken for process (a) in the case of the user with the OuterUser role was almost three times as long as for the user with the Doctor role while the time taken for process (a) was almost the same in both cases.

The reason why the time taken for process (b) was different depending on the user's roles can be explained as follows. As described in Section 4, either the Java method of getDoctorXSLT(), getInnerXSLT() or getOuterXSLT() is invoked depending on the user's role for filtering data. In our implementation of the Data filtering module, whether or not the user invokes such Java methods is sequentially checked by PERMIS. More technically, PERMIS checks whether to invoke a Java method whenever it is called. In this experiment, all of these three methods were checked in the case of the user with the InnerUser role, while only getDoctorXSLT() was checked for the user with the Doctor role. Therefore, the time for process (b) in the case of the user with the InnerUser role was three times as long as the process for the user with the Doctor role.

Table 2 System overhead.

	Total time	Process (a)	Process (b)
User (Doctor Role)	416 ms	149 ms	267 ms
User (OuterUser)	773 ms	457 ms	276 ms

From this result, it is easily predictable that the total time necessary for access control and data filtering functions will increase in proportion to the number of roles. For example, the maximum total time would be around 1,500 ms from a simple calculation if 10 roles were set up for the clinical database on this system. We consider this number acceptable but it will have impact on the practical use of the clinical database for Parkinson's disease research and diagnosis, since actual research and diagnosis requires more segmentalized roles such as attending doctor resident, nurse, and X-ray technologist. Furthermore, we have to consider the searching time of the XML database in addition to the number of the users. Therefore, we have to reduce the processing time through solutions such as having cache mechanism of attribute certificate and styleseet for role deployed in data filtering function.

5.2 Administration of Security Policy and Configuration

As described in Subsection 3.3, low-cost administration of security policy and configuration is an important factor for keeping the database secure. From this perspective, we discuss the administration of security policy and configuration in the access control mechanism using the cross-organizational environment shown in **Fig. 6**.

In this environment, we assume the following three things. Firstly, each organization of Hospital A, Research Laboratory B, and Hospital C agrees to use the role-model composed of Doctor, InnerUser and OuterUser and a specific clinical data format. Second, Alice, a user at Research Laboratory B, is working as a doctor at Hospital A and collaborates with a research group at Hospital C. Thus, she is assigned the Doctor role at Hospital A and OuterUser at Hospital C. Third, only Hospital A and Hospital C own clinical databases, each of which has a different data set and the proposed access control deployed.

Under this assumption, we first consider what the administrators at each organization have to do in case that Alice starts working as a medical doctor at Hospital C for some reason. In this case, the administrator of Hospital C has to

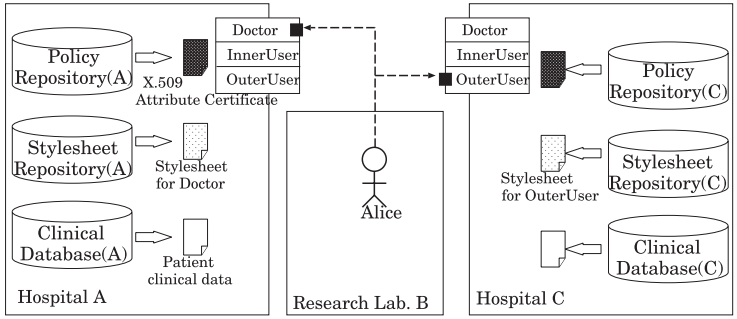

Fig. 6 A cross-organizational environment assumed for discussion.

change her role from OuterUser to Doctor immediately so that she has the privilege to browse the clinical data elements necessary for her diagnosis. Technically, all the administrator of Hospital C has to do is to revoke the X.509 attribute certificate assigning her to OuterUser and then issue the attribute certificate assigning her to Doctor as a new security policy. The administrative work here is performed independently of other involved organizations and no administrative work takes place at Hospital A and Research Laboratory B.

Next, we consider another case in which Hospital C adopts a looser site-security policy pertaining to data exposure to allow users with the InnerUser role to browse more security-sensitive data elements for some reason. In this case, all the administrator of Hospital C has to do is to re-register the stylefile for the InnerUser role to satisfy the new security policy. Importantly, the administrative work is performed only at Hospital C independently of other involved organizations and no work takes place at other organizations.

As discussed using two possible realistic cases, the proposed access control mechanism facilitates the administrators in responding to the changes of both the user's role derived from human real-world activities and the site-security policy pertaining to data exposure. However, an inevitable case that requires much administrative work for continuous operation may take places when the first assumption breaks down. For example, if a case should happen such that any of the three organizations want to use a role-model completely different from the current role-model for some reason, then all involved organizations would have to rebuild a policy repository and a stylesheet repository at each organization for continuous operation of clinical databases. Alternatively, the organization might have to give up sharing the clinical database with other involved organizations.

In the case shown in the above paragraph, the proposed access control is very weak. However, we consider that the cases where the first assumption breaks down happen less often than the cases where the user's role and security policy pertaining to data exposure change. Therefore, once we can succeed to build the realistic role-model and data format with which multiple organizations can agree, the administration of security policy and configuration for this access control mechanism is considered to be not labor-intensive and practical.

6. Related Works

Today, a couple of authorization technologies exist which are available for realizing the access control mechanism in combination with GSI. This section briefly reviews three possible alternative technologies of PERMIS, that is, CAS, VOMS, which are widely accepted and utilized in recent Grid research, and Akenti.

CAS (community authorization service) is an authorization service that allows the resource administrator to delegate a portion of privileges related to the resources, such as read and write operations to data resources, to the administrator of the Grid environment. With CAS, the authorization process is performed in the following way. In CAS authorization, the CAS server and database manages user group information and policies describing what kind of operations each group is permitted respectively. When the user accesses a resource of interest, s/he first has to contact the CAS server to ask for permission to access the resource. In turn, the CAS server returns information on what kind of operation s/he is permitted to perform (technically a GSI proxy credential with the information embedded). After that, the user can access the resource of interest with the credential but his/her access is further restricted based on local resource administration policy.

VOMS (Virtual Organization Management System) is a system for managing authorization data withing multi-institutional collaborations [4),20)]. VOMS allows the administrator of the Grid to manage user group information based on user role and attribute in a similar way to CAS. Therefore, the access to a resource

in the case of VOMS is performed in almost the same way as CAS. However and importantly, VOMS itself does not provide an authorization decision service, but a group management service for authorization. Therefore, the VOMS is often used for realizing access control together with authorization decision software such as LCAS (Local Centre Authorization Service) [21].

CAS and VOMS could be used instead of PERMIS as authorization technologies for the proposed access control mechanism. For example, in the case of CAS, if the administrator of the Grid manages a group of medical doctors, co-medicals and collaborators and then assigns a set of permitted operations to each group, and then the administrator of the clinical database permits the user to access the database based on local security policy through the filtering, the access control mechanism would be able to control the user's access depending on the user's group. As described in Ref. 22), however, CAS and VOMS do not provide the ability for the resource administrator to set the policy for access to his/her resources and then let the authorization infrastructure enforce the policy on his/her behalf. In other words, the final policy enforcement point in the case of CAS and VOMS is left to local resource administration settings and configuration, which may result in additional complex and error-prone configurations and settings. In contrast, as described in Subsection 4.2, PERMIS provides a robust functionality of Java-method-level invocation for policy enforcement as well as the X.509 attribute certificate-based policy decision functionality. This is one of the reasons why we have not adopted CAS and VOMS, but PERMIS, despite the majority and popularity of VOMS and CAS.

Akenti is an authorization infrastructure that allows for certificate-based access control for widely distributed resources [5]. It has a very similar architecture to PERMIS [23]. In fact, both PERMIS and Akenti commonly write their policies in XML and store them in certificates. Also, both offer policy decisions and policy enforcement functions based on PKI and PMI concepts based on their certificates.

Taking the architectural similarity between PERMIS and Akenti into consideration, Akenti can be leveraged instead of PERMIS for the access control mechanism proposed in this paper. However, Akenti uses its proprietary format for certificate description, while PERMIS X.509 uses a standard format. Furthermore, Akenti has adopted the traditional and well-known discretionary access control (DAC) scheme using an access control list, while PERMIS has adopted the recently-focused Role-based access control (RBAC) scheme which has emerged as the primary alternative to DAC and mandatory access control (MAC) because it is much better suited to the needs of commercial users than traditional schemes such as DAC and MAC [24]. In this research, after consideration of these two differences, we have adopted PERMIS rather than Akenti for the access control mechanism of the clinical database for Parkinson's disease research and diagnosis.

7. Conclusion

In this paper, we presented a fine-grained access control mechanism built into the clinical database for Parkinson's disease research and diagnosis. The access control mechanism can control the access to the clinical data on an XML data element level through the synergy of MyProxy, GSI, PERMIS and XSLT. This paper indicates that administration of our access control mechanism is not labor-intensive to the changes of users' roles and security policy related to data exposure but is labor-intensive to the change of data format of clinical data and role-model when sharing in a cross-organizational environment. Also, the overhead time incurred by the access control mechanism is shown to be acceptable but must be reduced for practical use.

A developed clinical database is now being operated experimentally in the Department of Neurology at Osaka University for the further evaluation of the practical usefulness and security of the clinical database, in the hope that the database can be deployed and utilized in the actual Parkinson's disease research and diagnosis. We would like to further improve the proposed access control mechanism by considering such issues as a practical access control model and incurred overhead time.

Acknowledgments This research is partly supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Young Scientists (B) [No. 20700062]. Also, this research was partly supported by Takeda Science Foundation.

References

1) Foster, I., Kesselman, C. and Tuecke, S.: The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *International Journal of High Performance Computing Applications*, Vol.15, No.3, pp.200–222 (2001).
2) Butler, R., Engert, D., Foster, I., Kesselman, C., Tuecke, S., Volmer, J. and Welch, V.: A National-Scale Authentication Infrastructure, *IEEE Computer*, Vol.33, No.12, pp.60–66 (2000).
3) Pearlman, L., Welch, V., Foster, I., Kesselman, C. and Tuecke, S.: A Community Authorization Service for Group Collaboration, *Proc. IEEE 3rd International Workshop on Policies for Distributed Systems and Networks*, pp.50–59 (2002).
4) Alfieri, R., Cecchini, R., Ciaschini, V., dell'Agnello, L., Frohner, Á., Gianoli, A., Lőrentey, K. and Spataro, F.: VOMS, An Authorization System for Virtual Organizations, *Proc. 1st European Across Grid Conference*, pp.33–40 (2003).
5) Thompson, M., Johnston, W., Mudumbai, S., Hoo, G., Jackson, K. and Essiar, A.: Certificate-based Access Control for Widely Distributed Resources, *Proc. 8th USENIX Security Symposium* (*Security '99*), pp.215–228 (1999).
6) Chadwick, D.W. and Otenko, A.: The PERMIS X.509 Role Based Privilege Management Infrastructure, *Future Generation Computer System*, Vol.19, No.2, pp.277–289 (2003).
7) Date, S., Tashiro, T., Nozaki, K., Nakamura, H., Sakoda, S. and Shimojo, S.: A Grid-ready Clinical Database for Parkinson's Disease Research and Diagnosis, *Proc. 20th IEEE International Symposium on Computer-Based Medical Systems*, pp.483–488 (2006).
8) National Parkinson Foundataion. http://parkinson.org
9) Araki, K., Ohashi, K., Yamazaki, S., Hirose, Y., Yamashita, Y., Yamamoto, R., Minagawa, K., Sakamoto, N. and Yoshihara, H.: Medical Markup Language (MML) for XML-based Hospital Information Interchange, *Journal of Medical Systems*, Vol.24, No.3, pp.195–211 (2000).
10) Dolin, R.H., Alschuler, L., Beebe, C., Biron, P.V., Boyer, S.L., Essin, D., Kimber, E., Lincoln, T. and Mattison, J.E.: The HL7 Clinical Document Architecture, *Journal of the American Medical Informatics Association*, Vol.8, No.6, pp.552–569 (2001).
11) Abramson, D., Lynch, A., et al.: Deploying Scientific Applications to the PRAGMA Grid Testbed: Strategies and Lessons, *Proc. IEEE International Symposium on Cluster Computing and the Grid* (*CCGrid2006*), pp.241–248 (2006).
12) HealthGrid Initiative. http://www.healthgrid.org
13) Biomedical Informatics Research Network (BIRN). http://www.nbirn.net
14) Pacific Rim Applications and Grid Middleware Assembly (PRAGMA). http://www.pragma-grid.net
15) Welch, V., Siebenlist, F., Foster, I., Bresnahan, J., Czajkowski, K., Gawor, J., Kesselman, C., Meder, S., Pearlman, L. and Tuecke, S.: Security for Grid Services, *Proc. 12nd International Symposium on High Performance Distributed Computing*, pp.48–57 (2003).
16) Foster, I., Kesselman, C., Tsudik, G. and Tuecke, S.: A Security Architecture for Computational Grids, *Proc. 5th ACM Conference on Computer and Communications Security Conference*, pp.83–92 (1998).
17) Chadwick, D., Otenko, A. and Ball, E.: Role-based Access Control with X.509 Attribute Certificates, *IEEE Internet Computing*, pp.62–69 (2003).
18) Basney, J., Humphrey, M. and Welch, V.: The MyProxy Online Credential Repository, *Software: Practice and Experience*, Vol.39, No.9, pp.801–816 (2005).
19) Tuecke, S., Welch, V., Engert, D., Pearlman, L. and Thompson, M.: Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile, *Internet Engineering Task Force*, RFC3820 (2004).
20) Alfieri, R., Cecchini, R., Ciaschini, V., dell'Agnello, L., Frohner, Á. Lőrentey, K. and Spataro, F.: From gridmap-file to VOMS: Managing Authorization in a Grid Environment, *Future Generation Computer Systems*, Vol.21, No.4, pp.549–558 (2005).
21) Steenbakkers, M.: Guide to LCAS verstion 1.1.16. http://www.dutchgrid.nl/DataGrid/wp4/lcas/edg-lcas-1.1/lcas.pdf (2003).
22) Chadwick, D.: Authorization in Grid computing, *Information Security Technical Report*, Vol.10, No.1, pp.33–40 (2005).
23) Chadwick, D. and Otenko, A.: A Comparison of the Akenti and PERMIS Authorization Infrastructures, *Proc. ITI 1st International Conference on Information and Communications Technology* (*ICICT 2003*), pp.5–26 (2003).
24) Ferraiolo, D.F., Kuhn, D.R. and Chandramouli, R.: *Role-Based Access Control*, Artech House Publishers (2003).

(Received July 18, 2008)
(Accepted August 8, 2008)
(Released November 28, 2008)

(Communicated by *Tatsuya Akutsu*)

Susumu Date received his B.E., M.E., and Ph.D. from Osaka University in 1997, 2000, and 2002, respectively. He was an Assistant Professor at the Graduate School of Information Science and Technology, Osaka University from 2002 to 2005. He also worked as a Visiting Scholar in University of California, San Diego in 2005. He worked as a Specially-appointed Associate Professor for the Internationalization of Education in the Graduate School of Information Science and Technology, Osaka University through the MEXT-funded educational program from 2005 to 2008. From 2008 he is working as an Associate Professor of the Cybermedia Center at Osaka University. His current research interests include application of Grid computing and related information technologies. He is a member of IEEE and IPSJ.

Kazunori Nozaki is a researcher with the Cybermedia Center, Osaka University. He received his D.D.S. from Hokkaido University in 2000, and his Ph.D. from Osaka University in 2004, respectively. He has been a Ph.D. student of the Graduate School of Information Science and Technology at Osaka University since 2006. His current research interests are in computational speech science. He is a member of IPSJ.

Haruki Nakamura is Professor of Protein Informatics at Institute for Protein Research, Osaka University. He received his B.S., M.A., and Ph.D. from the University of Tokyo in 1975, 1977, and 1980. His research field is Biophysics and Bioinformatics, and has so far developed several original algorithms in the computational analyses of protein electrostatic features and folding dynamics. He is also a head of PDBj (Protein Data Bank Japan) to manage and develop the protein structure database, collaborating with RCSB (Research Collaboratory for Structural Bioinformatics) in USA and MSD-EBI (Macromolecular Structure Database at the European Bioinformatics Institute) in EU.

Saburo Sakoda is Professor of Neurology at Osaka University Medical School. He received his Ph.D. from Osaka University in 1984. His research field is neuroimmunology and movement disorder in neurodegenerative diseases, and has so far developed a new drug to treat multiple sclerosis and a few systems to evaluate movement of fingers & muscular tonus. He is also the head of Clinical Research Center at Osaka University Hospital, and has worked on clinical database of Parkinson's disease.

Shinji Shimojo received the M.E. and Ph.D. degrees from Osaka University in 1983 and 1986, respectively. He was an Assistant Professor with the Department of Information and Computer Sciences, Faculty of Engineering Science at Osaka University from 1986, and an Associate Professor with Computation Center from 1991 to 1998. During this period, he also worked for a year as a Visiting Researcher at the University of California, Irvine. He has been a Professor with the Cybermedia Center (then the Computation Center) at Osaka University since 1998, and from 2005 and 2006 had been the director of the Center. He is an executive researcher in National Institute of Information and Communications Technology. His current research work is focusing on a wide variety of multimedia applications, peer-to-peer communication networks, ubiquitous network systems, Grid technologies and new generation network research. He was awarded the Osaka Science Prize in 2005. He is a member of IEEE, IPSJ and IEICE.