Open Social based Group Access Control Framework for e-Science Data Infrastructure

Hui Zhang, Wenjun Wu, ZhenAn Li
Beihang University, Beijing China
Outline

• Access Control Framework for e-Science Data Infrastructure

• Scientific Team Management based on Social Network

• Use Case - medical research
• **National scientific infrastructure to promote data sharing and collaborative researches**
China eScience Data Portal needs cross-domain Authorization

- A dozen domain-specific web portals
  Life science, environmental science, health care, climate research,
- Cross-domain authentication and authorization must be devised to enable scientists to have the single-sign-on service for unified data access.
1. Authorization model based on VO

- VO based Group Management and Access Control
  - PKI based authentication and authorization framework.
  - Two major VO management schemes: CAS and VOMS
- Issues:
  - Administrative burden of gridmap management
  - Heavy-weighted in ad-hoc collaboration scenarios
2. Attribute-based authorization

- SAML: an XML standard for exchanging authentication and authorization data between security domains.
- Through SAML and attribute-based access control, a scalable VO identity management can be developed by federating deployed campus identity systems.
3. Community Account Model

- Community Account Model in Science Gateways
  - a community VO identity for all the science gateway users
  - Allocated community resources for the account

- Attribute-based Authorization
  - SAML tokens carrying user attributes: name, ip address,...
  - Community credentials with SAML tokens
  - PDP & PEP (authorization server & gatekeeper) extended to check SAML tokens

- Not good enough for impromptu scientific collaboration
  - Community resource allocation from institution
  - Users can’t contribute their own resources
Motivation

• None of the frameworks can
  – Capture the social trustworthiness among researchers
  – Support social trust based resource sharing

• Scientists need
  – a group-oriented workspace to analyze, annotate and validate their datasets with different work styles and practices

• Goal of this research
  – Provide a scalable and flexible access control for sharing scientific data resources within research groups
Outline

• Access Control Framework for e-Science Data Infrastructure

• Scientific team management based on Social network

• Use Case - medical research
Multi-tenancy Access Control Model in the context of Social Trust

- Online Social Network based trust model
  - Social trust for scientific collaborations among researchers
- Friend-of-Friend (FOF)
  - Enables the description of transitive trusts
  - the depth of a trust relationship:
    - the length of the shortest paths between two people in the social graph
Multi-tenancy Access Control Model in the context of Social Trust

• **Scientific Team:** a 5-tuple $ST= <O,M,GC,RL,RS>$
  
  – $O$ represents the initiator of this team,
  
  – $M$ is the set of all the team members qualified by Friend-of-Friend rules in $GC$
  
  – $GC$ is the rule set issued by $O$, specifying the criteria for establishing this group
  
  – $RL$ is a resource list that includes all the resources visible to the members
  
  – $RS$ is the set of the access control rules issued by the group members for their resources.
Multi-tenancy Access Control Model in the context of Social Trust

• A multi-tenancy authorization model
  – <Issuer, Resource, Subject, Action, Condition>
  – The Issuer field enables every user to define his authorization policy over the resource objects he owns.
  – The field Resource are defined in the format of URI with the scheme and the resource path to describe RESTful Cloud services.
A multi-tenancy authorization model

• The field of *Subject* is defined as a 2-tuple:
  
  \(<\text{role}, \{\text{User}_1, \text{User}_2, \ldots \text{User}_n\}\>

• The field *Action* describes which operation is allowed to access the resource object

• The field *Condition* defines logic constraints for the action to be permitted in accordance with security policy
OAuth2.0 based Group Access Control Framework
OAuth

• Address the authorization delegation problem for web mashup sites
  – Third-party web sites want to access a user’s web resources

• Three-legged protocol steps
  – The Client obtains an unauthorized Request Token from the Authorization Server.
  – The User authorizes the Request Token through the Authorization Server.
  – The Client exchanges the Request Token with the Provider for an Access Token
OAuth 2.0

• OAuth 2.0 is the next evolution of the OAuth protocol
• Greatly extend the client profiles
  – Authorization flows for web browsers, desktop applications and smart phones
• Introduced a long-lived refresh token
  – Renew an access token with limited lifetime
• Widely supported
  – GoogleDoc, Microsoft Azure, Amazon EC2/ S3, Facebook
OAuth Protocol flow

• Three-legged protocol steps
  • The client obtains an unauthorized Request Token from the Authorization Server
  • The User authorizes the Request Token through the Authorization Server
  • The Client exchanges the Request Token with the Provider for an Access Token
Open Social based Group Access Control Frameworks
How it works

• The group authorization manager (PDP)
  – Stores the group membership, user-defined access control policy and the virtual resource pool

• Step 1 through Step 3
  – Resource owner registers their resources
  – Delegate the authorization to the group authorization manager

• Step 4-5
  – Users make request for accessing the resources
  – Granted by the manager with the access token
Resource Management and Access Control

• OAuth based vs PKI based
  – Address the authorization delegation problem for web mashup sites
  – Third-party web sites want to access a user’s web resources
    – Supported by Google Doc&Data, Azure, Facebook

• User contributed data resources
Comparison among three authorization solutions

<table>
<thead>
<tr>
<th></th>
<th>VO based authorization</th>
<th>Community Account Model</th>
<th>Social group model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>PKI infrastructure</td>
<td>PKI, SAML</td>
<td>OpenSocial, OAuth</td>
</tr>
<tr>
<td>Scalability</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Ad-hoc collaboration support</td>
<td>No</td>
<td>Partially</td>
<td>Fully</td>
</tr>
</tbody>
</table>
Outline

• Access Control Framework for e-Science Data Infrastructure
• Scientific team management based on Social network
• Use Case - medical research
**USE CASE – Medical Research**

- Research on connection between infants defect with parents conditions when giving birth
  - Alice has many infants defect data
  - Bob owns a data mining Web service for analyzing medical data

- **The way to collaborate**
  - Both Alice and Bob register their resources
  - Both of them set their own resources access policy
### Alice’s resource list

<table>
<thead>
<tr>
<th>Index</th>
<th>Title</th>
<th>UploadTime</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1986 National Birth Defect data</td>
<td>2012-06-16</td>
<td>Delete, Detail</td>
</tr>
<tr>
<td>2</td>
<td>1987 National Birth Defect data</td>
<td>2012-06-16</td>
<td>Delete, Detail</td>
</tr>
<tr>
<td>3</td>
<td>1988 National Birth Defect data</td>
<td>2012-06-16</td>
<td>Delete, Detail</td>
</tr>
<tr>
<td>4</td>
<td>1989 National Birth Defect data</td>
<td>2012-06-16</td>
<td>Delete, Detail</td>
</tr>
<tr>
<td>5</td>
<td>1990 National Birth Defect data</td>
<td>2012-06-16</td>
<td>Delete, Detail</td>
</tr>
</tbody>
</table>

Each dataset is a collection of birth defect records, where each row consists of 118 attributes containing critical defects classified by ICD (International Classification of Diseases).
Access Control Rules for the team of Alice and Bob

- $O=\text{Alice}$, means the group creator,
- $M=\{\text{Alice,Bob}\}$, means the group member
- $GC=\{\text{Alice, Colleague, 1}\}$, means the group creating rule
- $RL=\{\text{“1986 Data”, ..., “1990 Data”, “AR Service”}\}$
- $RS =\{\text{Rule1}\}$, means all the resource policies in the group. $\text{Rule1 is defined as}$
  \{Bob, “view, download”, “Time between 9:00am and 5:00pm”\}
# USE CASE – medical research

## Web Interface to Birth Defect Data

### Association Rule Mining

- **Birth Defect Data Input**
- **Birth Defect Types Input**
- **Birth Defect Possible Causes**

**Input Data Records:**
- Malformations of trachea
- Malformations of larynx
- Malformations of heart
- Down's syndrome
- Cystic kidney disease
- Hypoplasia
- Hydramnion
- Hydrocephalus
- Spina bifida
- Microphthalmos
- Undescended testes
- Anencephaly
- Cystic kidney disease

**Possible Causes:**
- Maternal smoking
- Maternal alcohol consumption
- Maternal infection
- Maternal diabetes

**Mining Arguments:**
- Support: 50%
- Confidence: 60%

**Output:**
- Ancephaly

**Frequent Patterns:**
- Maternal smoking → Anencephaly
- Maternal alcohol consumption → Anencephaly

**Possible Inferences:**
- Maternal smoking and alcohol consumption increase the risk of birth defects.
USE CASE – medical research

- Alice’s editing & adding ACL policies interface
CONCLUSION

• Introduce a social trust based, multi-tenancy authorization model for cross-domain resource access control.

• Based on the model, we design a group authorization scheme using OAuth 2.0.

• The use case about the collaborative medical research on prediction of birth defects demonstrates the effectiveness of our framework.
Q & A