Using Ontologies to Mitigate LDAP Deficiencies

Joshua Powers
Securboration, Inc.
1050 W. Nasa Blvd. Suite 156
Melbourne, FL 32901
jpowers@securboration.com

Abstract. Semantic technology powered access control schemes have been recently proposed to enhance the flexibility of role-based access control (RBAC) and its variants. These access control mechanisms depend heavily on rich, contextual data sourced from an identity attribute store. Unfortunately, most identity stores in use today use the Lightweight Directory Access Protocol (LDAP) representational schema which has several deficiencies as a knowledge representation, particularly when applied to fine-grained, contextual access decision policies. This paper reviews some of these gaps and shows how the same semantic infrastructure used for the access control mechanisms can be employed to mitigate LDAP assumptions.

Keywords: access control, identity management, authorization, semantic technology, LDAP, RDF/OWL

1. Introduction

In the past decade, the defense and intelligence communities have acknowledged the importance of moving from a 'need to know' assumption to a 'need to share' assumption with respect to the secure exchange of information [1] [2]. This has been interpreted in a number of ways, including reducing barriers between networks, establishing enterprise service buses, and building metadata repositories, federated search schemes, enterprise catalogs and enterprise-level portals.

At the same time, adoption of Service Oriented Architecture (SOA) standards has made the information delivery mechanisms themselves increasingly modular and decoupled from stovepipe systems of record. Access to authoritative data about a subject of interest requires the availability of a simple endpoint, usually a URL over some standard protocol, rather than a complex point-to-point integration between two large networks or systems.

These changes have not gone unnoticed by the information assurance and security communities. Mandatory Access Control (MAC) schemes that protect data at different levels of classification are still largely in effect, although secure cross-domain technologies are attempting to break some of those sharing barriers. More importantly, within the same classification level, Discretionary Access Control

(DAC), or any variant involving assignment of individual requestor privileges to individual resources, cannot scale to a goal of ubiquitous information sharing with unanticipated but qualified requestors.

To address the sharing assumption, information assurance efforts have looked at Role-Based Access Control (RBAC) [3], a more flexible protection model initially developed for industry, and a more generic formulation called Attribute-Based Access Control (ABAC) [4]. This model's original characterization is fairly vague in terms of specifying representational mechanisms, so semantic technology approaches have been suggested for formalizing ABAC. While these access control models have advanced to keep up with new information sharing requirements, there is an unfortunate gap in the representational state of the authoritative data that provide the critical information about requestors used to decide and enforce policies under these advanced access control models. These data are most often stored and managed in Lightweight Directory Access Protocol (LDAP) directories. In this paper, we first describe a couple of semantic technology-based access control schemes and the underlying identity attributes they require. Then we show the specific technical barriers presented by LDAP in addressing these requirements. With each barrier, we show how semantic technologies similar to those used in the access control models and policies can be brought to bear to mitigate deficiencies in these attribute stores. We conclude the paper with suggestions for future work.

2. Semantic Access Control Schemes

RBAC itself does not limit the attributes associated with requestors to any particular degree of granularity, complexity or context. However, in practice, RBAC typically uses a Distinguished Name (DN) for identification purposes, plus a set of group memberships, role occupancies and basic demographic data. It does not usually account for attributes of entities which form a context around the requestor, the resource and the nature of the request.

One approach to increase the flexibility of an access control decision is the Semantic Policy Broker [5]. This mechanism causes authorizations to flow through an ontology, following its graph-like structure through an arbitrarily wide context. A natural language description of a complex policy might be:

> *"An engineer can view information about a mission which a piece of equipment that they work with supports if they are part of the organization that owns that mission."*

Under most interpretations of RBAC, this would result in a mapping of users to roles, each role representing their participation in a mission:

```
mission1_role
   roleOccupant: user1
   roleOccupant: user2
mission2_role
   roleOccupant: user1
```

. . .

All of the intermediate context involving membership in an organization, mission support, equipment, etc. is left to an administrator to work out role-by-role. Using the Semantic Policy Broker instead, an administrator translates such a policy into a SPARQL query such as:

```
(?requestor rdf:type sempbro:Person)
(?requestor sempbro:memberOf ?organization)
(?requestor sembpro:engineers ?equipment)
(?organization sempbro:owns ?mission)
(?equipment sempbro:supports ?mission)
```

This query then satisfies for some combinations of requestors and missions and does not for others.

Another approach is ROWLBAC [6], which represents the roles, requestors, resources and permission decisions of RBAC as OWL DL classes. Some attention is given to the temporal relevancy of roles, either determined by a requestor's own assertion or by some additional, higher-level rules regarding the different roles which are relevant to a given request type.

The authors of these approaches have carefully left the nature of the identity store which would support their rules with instance data out of the scope of their discussions. Organizations likely to benefit from advanced access control models such as those above are almost certain to have their identities stored and managed in an LDAP directory.

3. LDAP/LDIF

LDAP is a binary protocol for querying and modifying directory data. It also specifies the representational scheme which is used in these directories. This is serialized in readable text as LDAP Data Interchange Format (LDIF) entries of the following sort:

```
dn: cn=John Smith,ou=Users,ou=People,dc=dod,dc=mil,c=us
cn: John Smith
mail: jsmith@dod.mil
employeeid: 123456789
objectclass: top
objectclass: person
objectclass: organizationalPerson
objectclass: inetOrgPerson
```

There are two hierarchical representations of where John Smith resides in the directory: a sort of structural class membership given by the objectclass attributes, and a sort of group membership given by the distinguished name string. Objectclasses such as person, organization, and organizationalRole are predefined by various LDAP RFCs. They determine which attributes may be used in an entry

of that type. The distinguished name reveals the hierarchy of groups, each of which is an instance of one of these various objectclasses.

LDAP is well-suited to provide rapid lookup of simple attributes to determine who may join a network, use a printer or perform other basic functions. It is not particularly useful in representing the kind of contextual information needed for the advanced access control models discussed above.

4. Compositionality

A subject's LDAP unique ID within the directory is the concatenation of an entity's group memberships in inclusion order. This presents a fragility with respect to organizational change over time which LDAP administrators have recognized. As a result, almost no interesting group membership is asserted within a typical LDAP directory outside of basic 'User,' 'Admin,' and 'Roles'. These groups are then included within a high-level group representing the entire enterprise. This approach conflicts with access control schemes whose decisions are based on finer-grained group membership information. Within the DoD, it is common practice for each Department to set up a high level LDAP group for contractors, one for civilian employees, one for reserve duty members and another for active duty members. Many DoD contractors are in fact reserve members as well. This does not mean, from an LDAP perspective, that they have two roles with respect to the same organization. It means that they are actually two different people depending on which credential they present to an access decision point. A separation of unique identification from group membership statements is a natural approach in an OWL ontology:

```
<ldap:Person rdf:ID="Person1">
  <ldap:name>John Smith</ldap:name>
  <ldap:employeeid>123456789</ldap:employeeid>
  <ldap:memberOf rdf:resource="ldap#ReportingUnit12">
...
</ldap:Person>
```

A distinguished name string may be stored explicitly as another property in the ontology or may be constructed by a traversal of membership relations if it is needed for legacy purposes.

5. Transitivity

There are two types of properties in an LDAP structure: those which range over string values and those which range over distinguished names. Neither of these property types may enforce transitivity within the directory. Outside of the group memberships that make up the distinguished name structure and the structural objectclasses, there is no support for transitive properties. This means that any role

or permission based on such a property must be 'flattened out' representationally and added one-by-one for each 'level' of entity so connected by the property.

Within the DoD, there are transitive command properties that are critical for access decision making. Administrative Control (ADCON) is the military doctrinal interpretation of Federal government management responsibilities. Operational Control (OPCON) authorized the employment of resources to accomplish assigned missions. Tactical Control (TACON) authorizes direct control of movements or maneuvers.

OWL ontologies, and the reasoners that operate on them, have built-in support for transitive properties:

```
<owl:TransitiveProperty rdf:ID="ADCON">
  <rdfs:domain rdf:resource="#MilitaryUnit"/>
  <rdfs:range rdf:resource="#MilitaryUnit"/>
</owl:TransitiveProperty>
...
<ldap:MilitaryUnit rdf:ID="ReportingUnit12">
  <ldap:name>Tech Platoon 12</ldap:name>
  <ldap:ADCON rdf:resource="#ReportingUnit34">
...
</ldap:MilitaryUnit>
```

The basic LDAP directory hierarchies, both structural and group membership, may also be represented to support legacy uses of the data. However, for the advanced access control schemes discussed above, it is only necessary to represent the properties and classes dealing with those portions of the real world needed to make the access decision.

6. Administration

In an organization the size of the DoD, or even one of its Departments, managing thousands of roles across tens of thousands of units and associating them with millions of employees is a daunting task no matter what technology is used. Choosing a representational scheme that does not allow transitive properties and that concatenates unique IDs based on membership information that may change exacerbates the administrative issues.

More concerning for administrative complexity and resource use is that detailed access decisions do need to be made. If they are not supported by the LDAP infrastructure, which is usually at enterprise or sub-enterprise level, it becomes the responsibility of individual application administrators to put requestors on access control lists (ACL) for resources, one-by-one.

An enterprise-level attribute store which has the representational power to match the fine-grained access control needs of resources housed in disparate applications will reduce redundancy of administrative effort. It should also increase the

robustness of the organization's cyber defense posture since the distributed administrative burden makes it hard for an enterprise monitor to observe the actions of a single requestor across many applications.

7. Scalability

LDAP directories can be provisioned in distributed fashion, across a number of physical servers. However, the largest LDAP implementations generally cover a few million personal accounts with a couple dozen organizational accounts and a couple dozen attributes.

The Lehigh University Benchmark (LUBM) [7] is the open test platform for RDF/OWL triple stores. Triple stores regularly handle SPARQL queries over billions of triples on fairly modest servers [8] [9].

8. Discussion and Future Work

The implementation of the data store which an LDAP-compliant server uses is not specified by the protocol. All of the "ins and outs," however, must comply with the LDAP representational schema. This admirable decoupling offers the possibility of implementing an RDF triple store as the LDAP server's database and wrapping it with fully LDAP-compliant services. This would seemingly defeat the purpose of the triple store's more useful representational schema, but it would offer the possibility of a 'side-by-side' set of RDF/OWL and SPARQL services that could be used by the advanced access control schemes discussed above. Development of such a hybrid server will be part of our future work.

The Security Assertion Markup Language (SAML) is an important recent development for communicating authorization attributes. Its XML-based format currently assumes LDAP-like contents, but could be easily extended to allow direct reference to ontology assertions.

9. References

[1] J. X. Dempsey, "Moving from 'Need to Know' to 'Need to Share:' A Review of the 9-11 Commission's Recommendations". Testimony before the House Committee on Government Reform, August 3, 2004.

[2] Office of the Director of National Intelligence, Intelligence Community Directive Number 501. http://www.dni.gov/electronic_reading_room/ICD_501.pdf.

[3] D. F. Ferraiolo, D. R. Kuhn, "Role-Based Access Controls". In Proceedings. 15th National Computer Security Conference. 1992.

[4] H. Shen, F. Hong, "An Attribute-Based Access Control Model for Web Services". In Proceedings. 7th International Conference on Parallel and Distributed Computing, Applications and Technologies. 2006.

[5] B. McQueary, A. P. Stirtzinger, “Semantic Policy Broker Final Technical Report”. Prepared for Air Force Research Labs. 2009

[6] T. Finin, A. Joshi, et. al. “ROWLBAC - Representing Role Based Access Control in OWL”. In Proceedings. ACM Symposium on Access Control Models and Technologies. 2008.

[7] Y. Guo, Z. Pan, J. Heflin. “An Evaluation of Knowledge Base Systems for Large OWL Datasets”. In Proceedings. International Semantic Web Conference. 2004

[8] Ontotext LUBM Performance Report. http://www.ontotext.com/owlim/benchmarking/lubm.html

[9] O. Erling, I. Mikhailov. “RDF Support in the Virtuoso DBMS”. Technical Report. http://www.openlinksw.com/uda/wiki/OdbcRails/main/Main/VOSArticleRDF/rdfdb1.pdf