Shibboleth Attributes and Service Provider (RUcore) Requirements

1. Introduction

This document outlines the capability to be provided in R5.0 to provide for authentication and authorization capabilities using Shibboleth 2.0 and XACML. This document also summarizes the attributes and LDAP variables to be used to authorize access to video resources in R5.0.

2. R5.0 Capabilities and Required Modifications

The objective in R5.0 is to provide the underlying capability for authentication using Shibboleth and for authorization using xacml. However, it should be noted that all videos ingested into the R5.0 repository (either in the NJEDge or RUcore installation) will be publicly available. All Shibboleth/xacml related procedures will be invisible to the user of the R5.0 although we will need to test basic functionality to insure that we have the platform in place to develop the required user features in R5.1. Authentication and authorization will be enabled in R5.1 for licensed videos.

R5.0 will have an authorized access service provider function, common to the showfed and RUcore full record displays, for parsing any Fedora object's XACML restrictions before painting links to its datastreams. Users wishing to view restricted content will need to be authenticated at some point by the Shibboleth server, after which relevant information sufficient to determine access privileges will be stored by the service provider in session variables that will persist throughout their browsing sessions. Shibboleth authentication may take place through an initial login page in a special portal, or at any time when a user in a public portal such as "NJVid Commons" happens upon an object with restricted content. If the user has already been authenticated, the service provider will paint datastream links according to the restrictions indicated by an object's XACML policy. If a user has not yet been authenticated, instead of a link to the datastream in question there will be a link to a Shibboleth login page that will prompt for the user's information, create the necessary session variables, and redirect the user to the object's display page. If the user turns out to have sufficient privileges to access the restricted datastream, the service provider will now paint the link to that datastream; otherwise, it will display a message indicating that access is not permitted. Other than this behavior, which will only be triggered by an attempt to access an object with a restrictive XACML policy, there will not be any noticeable difference in the functionality of either showfed or the RUcore full record display. Shibboleth's Scoped Affiliation attribute (see next section) will provide enough information for us to implement this level of authenticated access in R5.0. Any XACML policies manually created for R5.0 will be limited to restricting access to one or more datastreams according to either userID or fedoraRole, or embargoing an ETD datastream based on a UTC date/time specification. The service provider will be self contained and essentially agnostic about both Shiboleth and XACML.

As an initial prototype, we have set up a mechanism translating Shibboleth-derived organizations, roles, and edu-person-principal-names (eppn) into a standard set of Fedora users that can be used to drive XACML policy distinctions. For instance, users like the following have been added to the fedora-users.xml file:

```
<user name="faculty-rutgers" password="faculty-rutgers">
    <attribute name="fedoraRole">
      <value>faculty</value>
    </attribute>
  </user>
<user name="student-rutgers" password="student-rutgers">
    <attribute name="fedoraRole">
      <value>student</value>
    </attribute>
  </user>
<user name="member-rutgers" password="member-rutgers">
    <attribute name="fedoraRole">
      <value>member</value>
    </attribute>
  </user>
<user name="faculty-njit" password="faculty-njit">
    <attribute name="fedoraRole">
      <value>faculty</value>
    </attribute>
  </user>
<user name="student-njit" password="student-njit">
    <attribute name="fedoraRole">
      <value>student</value>
    </attribute>
  </user>
<user name="member-njit" password="member-njit">
    <attribute name="fedoraRole">
      <value>member</value>
    </attribute>
  </user>
<user password="rjantz-rutgers" name="rjantz-rutgers">
  <attribute name="fedoraRole">
    <value>individual</value>
  </attribute>
 </user>
```

The fedora-users.xml file can be edited and updated at any time, and any changes take immediate effect. Though we expect this file to be relatively stable, we have put it under the control of the superuser administrator of dlr/EDIT.

Once users have been added to the fedora-users.xml file, they can now be targeted in XACML policies with conditions like the following:

```
<Condition FunctionId="urn:oasis:names:tc:xacml:1.0:function:not">
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-at-least-one-member-of">
<SubjectAttributeDesignator AttributeId="fedoraRole" MustBePresent="false"
DataType="http://www.w3.org/2001/XMLSchema#string"/>
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-bag">
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">administrator</AttributeValue>
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">faculty</AttributeValue>
</Apply>
</Apply>
</Condition>
```

allowing, more broadly, the admin and faculty from either Rutgers or NJIT, or more restrictively with the following:

```
<Condition FunctionId="urn:oasis:names:tc:xacml:1.0:function:not">
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-at-least-one-member-of">
<SubjectAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string"
AttributeId="urn:fedora:names:fedora:2.1:subject:loginId"/>
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-bag">
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">fedoraAdmin</AttributeValue>
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">faculty-rutgers</AttributeValue>
</Apply>
</Apply>
</Condition>
```

allowing the admin and only Rutgers faculty. Another very broad policy can be used to allow anyone from the Rutgers, NJIT, or other participating communities:

```
<Condition FunctionId="urn:oasis:names:tc:xacml:1.0:function:not">
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-at-least-one-member-of">
<SubjectAttributeDesignator AttributeId="fedoraRole" MustBePresent="false"
DataType="http://www.w3.org/2001/XMLSchema#string"/>
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-bag">
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">administrator</AttributeValue>
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">member</AttributeValue>
</Apply>
```

```
</Apply>
</Condition>
```

Or a policy can allow broad access to certain participants only:

```
<Condition FunctionId="urn:oasis:names:tc:xacml:1.0:function:not">
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-at-least-one-member-of">
<SubjectAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string"
AttributeId="urn:fedora:names:fedora:2.1:subject:loginId"/>
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-bag">
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">fedoraAdmin</AttributeValue>
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">member-rutgers</AttributeValue>
<AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">member-njit</AttributeValue>
</Apply>
</Apply>
</Condition>
```

A Shibboleth Scoped Affiliation is probed to seeing if it contains an attribute that can be permitted by a particular policy. An Affiliation containing a value of “faculty@rutgers.edu;member@rutgers.edu” is translated into the following array of usernames

@usernames = ”faculty-rutgers”, “member-rutgers”;

and then passed to Fedora through a simple REST URLs like the following:

```
foreach $usernames (@usernames) …
http://$username:$username@$FEDORAHOST:$PORT/fedora/get/$pid/$ds
```

The first username to succeed is returned with the appropriate link. If all possible usernames fail, an access denied link is returned instead.

Our test Shibboleth server is providing variables like the following from Rutgers and NJIT identity providers. Note the difference in the number of variables currently being provided.

Apache Environment for Rutgers

Shib-Application-ID default	
Shib-Session-ID	_95c12d118882788956bae402a36ba40a
Shib-Identity-Provider	https://shib.oirt.rutgers.edu/idp/shibboleth
Shib-Authentication-Instant	2008-10-29T21:00:53.107Z
Shib-AuthnContext-Decl	urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport
Affiliation	staff@rutgers.edu;alum@rutgers.edu;member@rutgers.edu;affiliate@rutgers.edu;employee@rutgers.edu
displayName	Jeffery A. Triggs
eppn	triggs@rutgers.edu
givenName	Jeffery
mail	TRIGGS@RUTGERS.EDU
o	Rutgers University
persistent-id	https://shib.oirt.rutgers.edu/idp/shibboleth!http://lefty64.sccnet.rutgers.edu/shibboleth!B/kpo2OVygtf2bx7gu9VzfmE0ak=

sn	Triggs
unscoped-affiliation	staff;alum;member;affiliate;employee

Apache Environment for NJIT

Shib-Application-ID default	
Shib-Session-ID	_330df394c75368c0bba516706290cccb
Shib-Identity-Provider	https://dill.njit.edu/idp/shibboleth
Shib-Authentication-Instant	2008-10-29T20:03:37.270Z
Shib-AuthnContext-Decl	urn:oasis:names:tc:SAML:2.0:ac:classes:unspecified
Affiliation	member@njit.edu;employee@njit.edu;staff@njit.edu;alum@njit.edu

3. Continuing Authentication and Authorization Control

Because of the way our links are painted allowing access to large datastreams through Apache rather than directly through the Fedora Access API, we need an integrated mechanism for controlling continuing access to restricted datastreams. Currently all public datastreams are served on demand as symbolic links through a temporary directory. In order to handle the continuing restriction of certain datastreams, the current service provider appends .htaccess "File" restrictions for the datastream resource links requiring Shibboleth authentication mirroring that used to gain authorization to access the resource.[1] For instance, any resource accessed by someone authenticated as a Rutgers staff member would be controlled by adding lines like the following to the .htaccess file in the temporary directory:

```
<Files rutgers-lib_12913-DJVU-1.djvu>
AuthType shibboleth
ShibRequestSetting requireSession 1
require affiliation ~ ^.*staff@rutgers.*$
</Files>
```

If the same resource was accessed by someone authenticated as an NJIT faculty member, a separate File restriction would be appended:

```
<Files rutgers-lib_12913-DJVU-1.djvu>
AuthType shibboleth
ShibRequestSetting requireSession 1
require affiliation ~ ^.*faculty@njit.*$
</Files>
```

A resource restricted to an individual user would require lines like these:

```
<Files rutgers-lib_22820-PDF-1.pdf>
AuthType shibboleth
```

[1] The .htaccess file is only altered in this manner when a new Shibbolized usertype is granted access to a restricted resource. This will keep the .htaccess file comfortably small throughout the lifetime of the temporary link cache.

ShibRequestSetting requireSession 1

require eppn ~ ^.*rjantz@rutgers.*$

</Files>

and so forth. Anyone accessing these resource links, which we typically cache between releases and thus persist for some time, would have to have the same Shibboleth affiliation or eppn as the user who initially gained access based on the Fedora XACML policy for that resource.

A similar effect could be achieved by creating the datastream resource links in subdirectories requiring Shibboleth authentication mirroring that used to gain authorization to access the resource. For instance, any resource accessed by someone authenticated as a Rutgers faculty member would be created as a symbolic link in a subdirectory named "rutgers-faculty" and controlled by broader .htaccess files like the following:

AuthType shibboleth

ShibRequestSetting requireSession 1

require affiliation ~ ^.*faculty@rutgers.*$

If the same resource were to be accessed by someone authenticated as an NJIT faculty member, a separate link to the resource would be created in a "njit-faculty" subdirectory with a different require line:

require affiliation ~ ^.*faculty@njit.*$

and so forth. Anyone accessing these resource links, which we typically cache between releases and thus persist for some time, would have to have the same Shibboleth affiliation as the user who initially gained access based on the Fedora XACML policy for that resource.[2] For the time being we plan to continue with the first approach outlined above while exploring the feasibility of a new method of streaming most datastreams outside the context of the Fedora API-A.

4. Attribute Overview

The information to be made available through attributes is collected and maintained by the home organization (an identity manager) and is stored in a user directory – LDAP or a compatible directory in another database. For NJVid, we will need to determine 1) if the following attributes are sufficient, 2) which attributes should be mandatory, and 3) how these attributes will map to local LDAP schemas and to a controlled vocabulary. The table below lists possible attributes. Note that these attributes have been recommended by C. Hedrick and S. Daniel as the minimum we need to get started. It is expected that we may only need "eduPersonScopedAffiliation" for R5.0 and R5.1 since this attribute will provide both the institution and the role.

[2] Note: the second approach would require a change of the current apache configuration, which restricts symbolic links to a single directory.

	Attribute	LDAP Name	Mandatory 5.1	Optional[3]
1.	Scoped Affiliation	eduPersonScopedAffiliation	X	
2.	Targeted ID	eduPersonTargetedID		X
3.	Surname	sn		X
4.	Given Name	givenName		X
5.	Email	mail		X
6.	eppn	eduPersonPrincipalName		X

[3] Optional attributes enable additional personal services beyond simple access authorization. Discussion is continuing, for instance, about authorization based on course registration or personal identity.

5. Attribute Descriptions

Scoped Affiliation	eduPersonScopedAffiliation
Description	Multiple values of the form *value@domain*, where *domain* is (typically) a DNS-like subdomain representing the organization or sub-organization of the affiliation (e.g., "faculty@rutgers.edu").
Semantics	Specifies the user's relationship(s) to the home organization
Permissible Values	faculty, student, staff, alum, member, affiliate, employee
Typical Usage	These values will appear in xacml policies and will be compared against attributes that are passed in through Shibboleth.
LDAP Syntax	Domain-qualified string
Example and Notes	faculty@rutgers.edu. Note that it is possible to have a string in which a sub-organization is represented, e.g. faculty@cs.rutgers.edu. C. Hedrick has indicated that it will be difficult them to provide the department or sub-organization.

Targeted ID	eduPersonTargetedID
Description	A single string value of no more than 256 characters that uniquely identifiers a user in an opaque, privacy-preserving fashion.
Semantics	This attribute offers a powerful alternative to the use of eduPersonPrincipalName (EPPN) as a user identifier within applications and databases. Its power lies in the fact that it offers a significant degree of privacy and control for users. It also tends to be more stable than EPPN because it doesn't change merely in response to superficial name changes.
Permissible Values	
Typical Usage	We will need to use targeted ID if we require an opaque user identity that will not likely change. Targeted ID might also be used for personalized services.
LDAP Syntax	A single string value of no more than 256 characters.
Example and Notes	C. Hedrick has recommended the use of eduPersonTargetedID, however initially he may have to provide uidnumber in lieu of targeted ID.

Surname	sn
Description	Multiple string values containing components of the users's "family" name or surname.
Semantics	
Possible Values	Smith, Jones, etc
Typical Usage	Used for display back to the user.
LDAP Syntax	String
Example	Jones

Given Name	givenName
Description	Multiple string values containing the part of the user's name that is not their surname or middle name.
Semantics	
Possible Values	Jack, Joe, Susan
Typical Usage	
LDAP Syntax	String
Example	Jack

Email	mail
Description	From RFC 4524: The 'mail' (rfc822mailbox) attribute type holds Internet mail addresses in Mailbox [RFC2821] form (e.g., user@example.com).
Semantics	Preferred address for the "to:" field of email to be sent to this person.
Possible Values	
Typical Usage	For alerting or feedback to a user.
LDAP Syntax	X.521(2001) "organization" object class
Example	user@example.com