Attribute identification & storage

Table of Contents

Attributes for Shibboleth

Once an identity provider (IdP) is set up and working the next stage of a Shibboleth project is to identify which attributes you want to use to make authorisation decisions. The use of attributes to make authorisation decisions is a recent development; names and values for attributes will evolve as service providers are set up and start to require them. The Internet2 eduPerson schema has been developed for this purpose.

About Attributes

An institution (or Identity Provider) is responsible for providing attributes about each of its members (e.g. staff in department x, student from department y, or the type of role of student or staff, as well as specific entitlement to restricted resources). The institution is also responsible for supplying an Attribute Release Policy so that administrators can choose which attributes should be released to which online resources (or Service Providers). Administrators may also become involved in deciding which attributes should be released to federations. Individual users also have the facility to override and institutional policy for the release of their own attributes so have control over their own privacy.

What are attributes?

Attributes are pieces of descriptive information about a user. They can theoretically be any piece of information about a user, however in the context of Shibboleth attributes are likely to be information which can be used to make access decisions, e.g. "user is member of staff" or "user is allowed to see medical content". Information, particularly personal information, such as "user is 137cm tall" is unlikely to be of use in terms of Shibboleth. Some attributes will describe a user's relationship to his home institution.

Why does Shibboleth need attributes?

Attributes are the key to the way Shibboleth authorises user access to resources. Shibboleth uses inform-

ation about a user to determine whether they will be granted access privileges. To illustrate this, currently the most common access decisions in use with Shibboleth would read:

"User A is a member of university X and can access journal Q because access university X bought access to journal Q.

User B is a member of university Y and can't access journal Q because access university Y hasn't bought access journal Q."

It is easy to envisage Shibboleth being used to facilitate much more complicated access decisions; i.e. a case in which only medical students taking the pathology course, who are enrolled for more than 2 years, can access autopsy photographs.

However, the Shibboleth Identity Provider software does not store or manage user attributes. To use the software effectively, an attribute store such as an LDAP directory or database needs to be set up. User data should be available on installation, or there should be some means of populating the data store set up.

Identifying useful attributes

Which individual attributes are chosen to make each authorisation decision will be determined by the following factors:

- the attributes that are required by the online application;
- the institution's privacy policy;
- which attributes can be collected from users in a timely and scalable manner.

Technical issues

Most of the currently contemplated attributes for use in U.K. and U.S. federations are simple and will probably present few technical issues to identity providers wishing to provide them. However when one begins contemplating attributes for internal use or attributes required for complex applications then technical concerns about how an identity provider would aggregate this information begin to emerge. While the use of most simple attributes will be determined by the interplay of application requirements and privacy concerns, technical concerns are going to determine what can and can’t be used as attributes in these more complex cases. There are several technical issues which may impact on the ability of an institute to use certain attributes. The first is the ability to obtain the attributes in a timely and scalable manner. The second is whether attribute are stored in a suitably structured store or stores.

Scalability

In order to use attribute with Shibboleth, the attribute aggregation process needs to be able to access and aggregate user attributes in a reasonable amount of time, i.e. seven seconds would be a reasonable figure to aim for. Therefore it is likely the the institution will need a reasonable and performant attribute store. Institutional grade stores like Microsoft Active directory, openLDAP, MSSQL sever, MYSQL, Postgress etc. would be suitable; whereas personal grade data stores like Excel or Access databases would not be suitable.

Dealing with complex storage structures

At present the facility for aggregating together attributes within Shibboleth is reasonably flexible and ro-

bust, however is not as flexible as may be required by some institutes. The current facility allows institutes to query both ldap and jdbc stores in order to obtain attributes. It is even possible to query both at the same time. However it is not possible to use conditional logic for obtaining attributes. This may be a problem where data about different user populations is held in different forms in different stores. For example if staff and student data are held in separate systems with different formats it would first be desirable to query to see if the user was staff of student then adjust the next query based on the reply in order to obtain the required attributes. If this is required, then a system to do this would have to be hand made. Shibboleth does support Java plugin classes to achieve this, however it would be a burden on that institute and would require someone with the requisite Java skills.

Another approach may be to try and decouple the attribute aggregation process from the Shibboleth infrastructure. If the Shibboleth identity provider were able to query a webservice in order to obtain attributes about a user then the webservice could be written on any platform in the language most suitable for querying the institution datastore. At present Shibboleth requires that the person who understands the structure and location of an institution's data knows Java in order to make attibutes available to the software. It would be better if this person could program in their language of choice (e..g. C# and VB for Microsoft based infrastructures, ABAP for SAP, etc) and then make the results available via a webservice. In many institutions the attribute aggregation process is a valuable piece of business logic which it would be desirable to reuse.

Management issues

The process of managing attributes is fundamental to the successful implementation of Shibboleth. It may well be reasonable to allow individual users to set some attributes for themselves yet be prevented from setting others, e.g. those determining access to national services. It would be easy to set up a separate directory to hold these attributes, but for a properly scalable solution this attribute store will need to be tied in very firmly with institutional databases.

The Attribute Management in an Institutional Environment (AMIE)[1] project has predicted that institutions will be faced with dealing with at least four different environments in which attributes for users must be held and managed. These attributes will control:

- access to national facilities: in most cases the attributes needed are relatively simple there are significant exceptions. At first glance the required attribute for a site licence is simply that the person is a ‘member’ of the institution. However, there are already national services which either restrict access to known individuals (Ordnance Survey) or restrict access to a particular group of users (medics);

- access to control access to (mainly) learning environments in small scale inter-institution collaboration: Institutions are increasingly collaborating with each other; the generic issues of who defines attributes, the ability to set attributes and trust needs to be established. Although the collaboration may be at the level of department, the level of authentication will lie at an institutional level in order to tie up who belongs to which institution;

- access to resources of Virtual Organisations: Although the issues associated with this aspect appear similar to the previous item, the ownership of the right to define attributes lies with the ‘virtual organisation’ rather than the union of the institutions concerned. These issues are the same as those governing the management of federations;

- access to locally held resources that need to be constrained to a particular group: There is a huge demand in institutions to control access for a set of people who do not fit neatly into a tidy organisational area. Shibboleth has the potential to enable this in a simple, scalable way but the mechanisms need investigation and documentation.

Although there is clearly some overlap across the above groupings, the ability to define the attribute, to

[1] http://www.ucs.ed.ac.uk/projects/amie/docs/AMIE_Project_Plan_Aug_04.doc

assign it, and the trust patterns needed to sustain the activities are different in each case.

Managers need to be aware what information is actually available from institutional management information systems, as this will be key to populating the attribute store that Shibboleth will use. Since different Service Providers may require different attributes, there will be a core set of attributes that it is necessary to store for each individual. Rights to keep data must be ascertained, though since holding this type of data is key to the institution fulfilling its contract to the staff and students, these rights are granted.

It is worth noting that, in the case of a Shibboleth connection, no personal data which comes under the control of the Data Protection act should be exchanged, therefore, the act does not apply. Shibboleth also gives privacy control on release of attributes. An Identity Provider would contract with the Federation that it will release data as necessary for authentication and authorisation processes; Service Providers will be contracted to only request necessary data about individuals in order to authorise use of the resources they protect. In many cases all data will be anonymised, e.g. in the case of an institutional subscription, the only data needed to authenticate individuals is "affiliation to university x".

The JISC Data Protection Code of Practice for the HE and FE Sectors provides guidance on best practice concerning the retention of records containing personal data (version 2 is available at: http://www.jisc.ac.uk/index.cfm?name=pub_dpacop_0101).

Attribute standards

The drive behind federated authentication is to be able to deploy and use services outside of a single institution. Therefore, there is a need for agreement over which attributes should be used and what format they should take. While Shibboleth can technically support each institution using different attributes and formats of data, this should not be intended and is not practical. Shibboleth is designed to be used in loosely coupled federations. The purpose of the federation is to decide on standards for attribute sharing.

In practice it is likely that the federations themselves will seek to standardise attribute formats and types across federations. The technical and administrative overhead of accessing different data about a user and distributing it in various formats mean that there is a large incentive to use global standards. One such standard is the eduPerson standard, which has come from the US educational sector.

What is eduPerson?

http://www.educause.edu/eduperson/

EduPerson was originally conceived, by an EDUCAUSE/Internet2 working group, as a schema for LDAP directories to store information which represents users in a standard way. It was envisaged that software used in one university could be reused in another with minimal need for recoding. It would also help commercial software suppliers develop systems for campuses. The eduPerson standard defines many different types of data that can be stored about a user. A subset of these data could prove useful in making authorisation decisions about a user.

What is UKeduPerson?

The UKeduPerson project[2] followed three strands: an assessment of the international picture; a “bottom-up” assessment of potential requirements for a UKeduPerson schema; and, a consultation exercise with Shibboleth-aware information vendors. The project resulted in the production of the UKeduPerson schema and made recommendations to JISC for future development. There has been no real impetus for internationalising the schema[3].

[2] UKeduPerson: www.angel.ac.uk/UKeduPerson/
[3] John PAschoud (LSE); UK EduPerson Project manager; in conversation, May 2005

Which attributes should I choose?

The use of attributes in the context of Shibboleth is an academic community driven process; it is therefore likely to be subject to "fashions". If the federations you wish to join and the services you wish to use require certain attributes then your institution will have to provide them in order to gain access to those services. That said, it is against both the federations and the service providers best interests to dictate too onerous a set of attribute requirements.

In practice a very limited set of attributes is currently being used. Only a slightly more expansive list is being contemplated for future developments by most federations.

The key currently used attributes are:

- eduPersonScopedAffiliation: used for the basic authorisation decision: does uni.ac.uk subscribe to the service?
- eduPersonPrincipalName: identifies an individual editing account
- eduPersonTargetedID: Many services can make use of, but do not require, the eduPersonTargetedID attribute. This is a persistent opaque identifier, which can be used to represent a user on a particular site but which is not used with other sites and enables service personalisation (remembering data such as stored searches about a user over different login sessions) without the service provider knowing the user's real identity.

Further attributes under consideration for use in federations (e.g. SDSS[4] in the UK) are:

- given name
- surname
- common name
- eduPersonEntitlement

The eduPersonEntitlement attribute is where all the individual entitlements (or capabilities) are held. In the SDSS project at Edinburgh University, the attributes currently supported are those deemed 'highly recommended' by the InCommon Federation which are those given in the two lists above, plus:

- email
- userid (uid)

The uid attribute can also be described as a Universal User Name (UUN). The uid is usually automatically assigned to staff and students at an institution, based on name or matriculation number, and can be assigned to registered visitors (including alumni) with format dependent on the category.

It is worth noting that some sensitive personal data will be automatically created through use of systems. Tags for (e.g.) reading habits are kept and in the real world could be used as personal identifiers. Fundamentally this type of data should not be collected, but it could be used to enhance the user's experience of using a resource. Anonymised personalised services can be offered through the use of the uid. Universities need to implement an attribute release policy to assure that attribute authority access is the only

[4] http://sdss.ac.uk/

way to get third party access to personal data. The attribute release policy needs to be machine readable; current institutional processes may not be secure enough, and ad hoc releases of data may fall under the Data Protection Act[5].

Which attributes does the federation need?

A common misconception is that the Federation defines what attributes can be exchanged. In fact federations can provide attribute flexibility, since the federation should only provide a framework for attribute exchange within a common set of policies. Which attributes are exchanged are a matter of agreement between the identity and service providers.

There are at least two ways to approach the issue

- to define all that users might want to access and attibutes to go with those items; however, without use cases this is tricky, and it should not be a case of what can be invented, but most usefully what can be done to solve the current issues
- to invent the minimum number of attributes, and use existing ones to solve use cases that already have issues (such as the sharing of medical resources being addressed by IAMSECT)

The latter option has the advantage of increasing interoperability when new use cases arise. An attribute allowing "different levels of entitlement" needs to be demonstrated.

See the IAMSECT document 'Joining a Shibboleth Federation' for more information on attributes and federations.

Storing and retrieving attributes

Storage solutions will tend to be institution-specific. A common choice is to use the Windows Active Directory as an attribute source. Both the IAMSECT project at Newcastle University and the SECURE project at LSE have identified that the Active Directory may not contain enough information. Also, if the Active Directory doesn't already contain the requisite information, trying to add it is difficult. There is more information regarding this in other IAMSECT deliverables and the SECURe project documentation directory[6]

. If another solution is sought, then initially it will necessary to identify a suitable institutional data feed or feeds.

The Attribute Authority (AA) is a service at the Identity Provider that is responsible for sending attributes associated with a user, to a trusted 3rd-party. The AA applies an Attribute Release Policy (ARP) to attributes that are made available to target resources. The ARP defines which specific information is released about a user. At the service provider, the SHIRE (SHibboleth Index Reference Establisher) component interacts with the SHAR (SHibboleth Attribute Requester). The SHAR requests attributes about the user, directly from the AA. A decision is then made by the SHAR (and the notional concept of a ‘Resource Manager’), based upon the requested attributes as to whether the user is authorised to access the resource. Although authentication is not actually part of the Shibboleth specification, it is inherently part of the authorisation process that the user’s identity is verified by their local institution. The mechanism for doing this is left up to the institution itself to define and operate.

[5]Additional guidance on the Data Protection Act and other legal issues surrounding the use of information is available from The Information Commissioners Office: http://www.informationcommissioner.gov.uk/eventual.aspx?id=34
[6]http://www.angel.ac.uk/SECURe/deliverables/documentation/directory.html

Aggregation of attributes

Once the attributes have been collected by the identity provider, they need to be aggregated.

Aggregated attributes can be used to provide a user profile. A user's identity may be represented as an aggregate of all his/her attributes, or aggregates of subsets of attributes, such as those associated with a user's professional identity as distinguished from his/her personal identity. No single attribute should identify a user on its own.

Shibboleth has a particularly strong focus on maintaining user privacy and controlling the release of user-specific information (such as personal attributes) to service provider's external to the user's organisation. Aggregation should only happen at the origin not on the target at the service provider; this should be a part of a federation's policies. Although the need for user privacy can been expressed in a policy, such policies may be vulnerable to being bypassed by data aggregation. This is an extension of the problem of propagating trust via delegation to a federation; i.e. how does a user meaningfully restrict and control the use of data in services that are invoked only indirectly?

Institutions need to have the right to collect and aggregate data in this way. There is no reason why all of anyone's personal information should be held on one server, and many reasons why it should not be. However, when the person wants to access their information, and control other people's access to it, it is important that they can have access to all of its parts, and access it in an easily-understood structure.

Release of attributes - who can?...

The ability to enable users to control the selective release of attributes to targets is an inherent property of the Shibboleth model. The extent to which this is a requirement for simple inter- and intra-institutional use, and the interaction of institutional policy and user choice will be investigated by the AMIE[7] Project at Edinburgh University. Possible mechanisms to control how users will specify release policy choices will also be considered during the SDSS project at Edinburgh.

[7] Attribute Management in an Institutional Environment (AMIE) http://www.ucs.ed.ac.uk/projects/amie/