Practical access to electronic journals via Shibboleth

Table of Contents

Introduction

Purpose

This document aims to set out the requirements for using Shibboleth in its simplest context. Access to electronic journals does not require complex authorisation decisions – in general only an assertion that the accessor is affiliated to a particular institution is needed – and so the issues of authorisation attribute storage and aggregation, and the complexities they entail, do not arise.

About Shibboleth

Shibboleth is a distributed authorisation project that has been developed by a federation of higher education establishments in the United States called Internet2. Shibboleth is not an authentication or authorisation scheme. It is an open, standards-based protocol for securely transferring attributes between an identity provider (local institution) site and service provider (resources) site which is supplied as an open-source reference software implementation.

About Access Control

Academic libraries licence and offer access to a range of online resources that are regarded as necessary to support research and teaching in any subject. One of the big issues to be solved is maintaining a bal-

ance between adhering to the legal and contractual responsibilities to publishers (to limit access to only those users covered by licence terms) and to the library's users (to keep personal information secure), and, finally, to ensure the fundamental function of a library operates - that users are shown the simplest path to the information they want[1].

About e-journals

E-journals are journals in either full or partial text found online; in academic institutions access to these is usually managed by a library. Most journals require a subscription, but some are accessible at no cost. Some e-journals are found only electronically; a large number of titles are published in both paper and electronic versions. E-journal providers can be divided into two categories: publishers and aggregators.

Most e-journal publishers provide access to their titles directly from their websites; some journal publishers make their e-journals available only on their own sites (e.g. Royal Society of Chemistry[2]).

Aggregator services offer access to e-journals from publishers that they have negotiated with. They will organise e-journals access, and administer passwords, IP address access, table of contents services, usage statistics and archiving. Moreover, users benefit from using a single interface. Examples of these services are SwetsNet[3] and Ebsco Online[4]. There are also companies which deal specifically with e-journals access, such as Ingenta[5].

Skills Set required

In order to Install and manage a Shibboleth software implementation you will need to be able to access the following skills:

- A reasonable working knowledge of the Linux (or Unix) Command Line Interface (CLI);
- Knowledge of how to use the apache web server, either the 2.0.* or 1.3.* versions;
- Familiarity with the concepts of https communication (certificates, keys and the like);
- Familiarity with firewalls or access to someone who is familiar, in particular with Linux iptables style firewalls;
- Familiarity with the setup of Windows Active directory, or access to someone who has those skills;
- Importantly, a willingness to read around subject areas, management pages and mailing lists, since this is an emerging and swiftly developing area.

Planning

There are several essential elements that must be present in the environment to ensure Shibboleth functions well (see section 2). Shibboleth is entirely written in Java on the Identity Provider side. The basic installation of the Shibboleth Identity Provider should be carried out before joining a federation, as the latter will test that the installation works, and will require information that will not be available until installation is complete.

[1] For additional information see Paschoud, J. (2005) Shibboleth and SAML: at last, a viable global standard for resource access management. *New Review of Information Networking* Vol. 10, No. 2. (November 2004), pp. 147-160
[2] http://www.rsc.org/is/journals/j1.htm
[3] http://www.swetsnet.nl/
[4] http://www-uk.ebsco.com/home/
[5] http://www.ingentaconnect.com/

Components of the Shibboleth system

Figure 1. Simplified components of a Shibboleth system

The process is as follows:

1. User requests resource
2. Request is redirected to WAYF (which can be hosted by the Federation)
3. The user selects their home institution from the list provided by the WAYF, which then redirects the request to the Identity Provider at that institution
4. The user authenticates at their home insitution via the WebISO (see below)
5. If authentication is successful a handle (digitally signed certificate) is passed to the Service Provider
6. The handle goes back to the Identity Provider to request attributes for authorisation
7. The requested attributes are returned to the Service Provider
8. If the attributes are acceptable, then access to the resource is authorised

The Identity Provider deals with authentication and the Service Provider has responsibility for authorisation. Shibboleth interactions are shown in full in Figure 2 on page 6 of the SWITCH publication *AAI System and Interface Specification*, available from http://www.switch.ch/aai/docs/AAI_System_Specs.pdf

Local Requirements

The following four requirements have to be met in order for users to be able to gain access to e-journals via a Shibboleth connection in its simplest context.

A Web-based institutional single sign on (WebISO)

Shibboleth requires a local method to authenticate users. Single Sign-On allows users to provide their username and password once to a trusted service and to have their identity securely, consistently and seamlessly provided to many other web applications.

An open-source WebISO solution, Pubcookie, can be obtained from http://www.pubcookie.org/, which also has documentation to describe the installation process. An installation guide detailing set up on Redhat AS 3.0 with authentication against Windows Active Directory is available from http://iamsect.ncl.ac.uk/deliverables/. The minimum requirement is for the WebISO to be able to authenticate browser users and supply their identity to the Handle Service (see section 2.3 below).

The Pubcookie login server has to be a standalone dedicated secure web server that only serves pubcookie login requests; it should not be used for other secure web serving or for other tasks. The server will form the main gateway for web-based login, so it is imperative that it is as secure as possible. It should therefore have as few applications running as possible in order to reduce the number of potential exploits.

Newcastle University has chosen to use Pubcookie, but Yale CAS[6] is also common as it comes as a

component of UPortal. The LSE uses Yale CAS and as part of the SECURe deliverables an installation guide, available at: http://www.angel.ac.uk/SECURe/deliverables/documentation/directory.html, was published.

The WebISO uses local directory services (see section 2.2 below) to authenticate the user, challenging for username, password, digital certificate, smartcard, biometric identification, or whatever the institution routinely accepts for authentication. However, this only occurs if this is the first authentication request of the session, thus single sign on is achieved.

Related technologies

Kerberos is another network authentication protocol. It is designed to provide authentication for client/ server applications by using secret-key cryptography. An open source implementation of this protocol is available from the Massachusetts Institute of Technology (MIT[7]). The Kerberos protocol uses strong cryptography so that a client can prove its identity to a server (and vice versa) across an insecure network connection. Some reports state that 'Kerberos can already complement or enhance the deployment of Shibboleth ... standards'[8]

A Local Directory Service

Shibboleth also requires a source of authentication from a common institutional directory service. Many institutions will already be using the Windows Active Directory for user authentication and Pubcookie or Yale CAS provide a flexible layer on top to provide web based secure single sign on.

Windows Active Directory directory service is a distributed directory service that is included with Microsoft® Windows Server 2003 and Microsoft® Windows 2000 Server operating systems[9]. A directory service provides a centralised location to store information in a distributed environment about networked devices and the people who use them. A directory service also implements the services that make this information available to users, computers, and applications. The Active Directory service is both a database storage system (attribute store) and a set of services that provide the means to securely add, modify, delete, and locate data in the attribute store. The minimum requirement is that user IDs need to be mapped to user attributes.

The most common alternative directory service is the LDAP (Lightweight Directory Access Protocol) directory, typically used to store information about entities such as people, offices, machines, but can be used to store information about anything described by a set of attributes. Being a protocol, it can be implemented in any way, as long as the protocol is adhered to. There are a number of implementations to choose from, including the open source OpenLDAP[10] from the University of Michigan; Oracle's[11] LDAP server that runs on top of an Oracle database. Sun One[12] (previously known as iPlanet) and the Novell LDAP services for e-Directory[13] are two more examples of software running the LDAP protocol.

A Shibboleth Identity Provider

Shibboleth has a complex architecture with a large number of components, but can basically be divided into two: the Identity Provider (origin) and the Service Provider (target). The Service Provider protects the resource, while the Identity Provider ensures that users are authenticated and passes on their attributes to the Service Provider on request.

The Shibboleth Identity Provider has two components, the handle service (HS) and the attribute authority (AA) (see Figure 1). The handle service is the component of Shibboleth through which the user au-

6 http://tp.its.yale.edu/tiki/tiki-index.php?page=CentralAuthenticationService
7 http://web.mit.edu/kerberos/
8 http://www.computerworld.com/printthis/2005/0,4814,100021,00.html
9 http://www.microsoft.com/technet/prodtechnol/windowsserver2003/library/TechRef/6f8a7c80-45fc-4916-80d9-16e6d46241f9.mspx
10 http://www.openldap.org/
11 http://www.oracle-base.com/articles/9i/OracleInternetDirectory9i.php
12 http://www.sun.qassociates.co.uk/software-solaris-9-dv.htm
13 http://www.nldap.com/NLDAP/

thenticates. Typically, they will be redirected to it by a Shibboleth Service Provider via a WAYF ('where-are-you-from' servicZetoc Searche) provided by an institution or a federation of institutions of which the user's institution is a member. The AA retrieves information about the user from the institutional directory system (the source of authentication mentioned in 2.2) or other attribute store, and forwards these to the Service Provider according to the identity provider attribute release policy.

In the Shibboleth architecture, the identity provider must have an X.509 server certificate for use in authenticating the servers to each other. This should be issued by a commonly accepted Certification Authority (such as Globalsign[14]), which may be stipulated in Federation rules. In theory, Shibboleth can accommodate certificates from any trusted source (e.g. a commercial CA or an ad hoc institutional CA). It is important to note that, when setting up the identity provider, you need to know which CA your chosen federation(s) will accept. In practice, management of a federation will be easier if all institutions get certificates from the same trusted source. For example, in the SDSS Federation[15] (see 2.4.1), this source is either Globalsign or, for development sites, the SDSS internal certificate service.

A guide to installing Shibboleth Identity Provider on Redhat AS 3.0 is provided as part of the IAMSECT deliverables at http://iamsect.ncl.ac.uk/deliverables/.

Join a Federation

A federation is a group of institutions and resource/ service providers who have common policies (often expressed in a trust agreement) which allow them to access the same online resources. In so doing, they must implicitly or explicitly agree to a common set of guidelines.

In a Higher Education Federation, users will usually come from identity provider institutions, which will host the Shibboleth identity provider server and supply the source of authentication. In the case of access to e-journals, the federation also needs to include the journals required by the users as service providers.

Joining a federation is not necessary for the deployment of Shibboleth, but joining one will dramatically increase the number of resource providers available to users without having to enter into a bilateral agreement with each one.

There are examples of production federations in Finland, Switzerland and the USA, see *An introduction to Shibboleth Federations* at http://iamsect.ncl.ac.uk/deliverables/ for more detail.

SDSS Federation (UK)

The SDSS Federation (www.sdss.ac.uk [???]) has several resources set up for a Shibboleth Identity Provider to target; these currently include the EDINA resources BIOSIS (life sciences), EMOL (film and video), UPDATE (farming, environment) and the Education Image Gallery plus the MIMAS hosted Landmap (satellite image and digital elevation data for the British Isles) and Zetoc Search (British Library Electronic Table of Contents), and the Internet2 Shibboleth Wiki. The federation also contains 7 identity providers and 8 JISC Core Middleware projects, including AMIE, SDSS, SPIE.

Once the prerequisites of installing an Identity Provider and obtaining an X.509 certificate have been met, there is a simple email form to complete to apply to join the SDSS federation as an identity provider (origin site)[16]. SDSS Wiki Documentation[17] explains how to configure the Shibboleth Identity Provider to take part in the federation. This document also describes the steps an identity provider needs to take to generate the eduPersonScopedAffiliation attribute automatically (without an attribute store) by setting the required scope in resolver.xml.

In the UK it would be practical for institutions wishing to complete the Shibboleth installation process in the near future to join the SDSS Federation. The current administration of SDSS is scalable to all UK in-

[14] http://www.globalsign.net/
[15] http://sdss.ac.uk/
[16] http://sdss.ac.uk/wiki/wiki.pl?JoinFedOrigin
[17] http://sdss.ac.uk/wiki/wiki.pl?SetupIdentityProvider

stitutions. SDSS is not a production federation (see 2.4.2), instead it offers a robust enough trust system for delivery of licensed content, but with fairly low entry requirements to encourage the participation of development projects.

Currently:

- all members of the federation are required to observe best practice in the handling and use of digital certificates and private keys.
- all identity providers (origins) must make reasonable attempts to ensure that only members of their institution are provided with credentials permitting authentication to the handle server, and that the assertions made to service providers by the attribute authority are correct.
- all service providers (targets) must agree not to aggregate, or disclose to other parties, attributes supplied by identity providers.

UK Production Federation

This issue is currently (June 2005) out to consultation in the interested community. Practical and policy issues underpinning the formation of a single production Shibboleth federation in the UK HE and FE communities have been outlined in a blueprint paper. An advisory board has also been set up to recommend the appropriate action to be taken by JISC to set up the federation following the conclusion of the consultation process. The aim is to write a set of drafts for more detailed policy papers by July 2005; realisation of these plans should occur in Autumn 2005.

Organisations (such as the early adopter projects and IAMSECT) currently using one of the pilot federations, run by Eduserv or SDSS, would be able to transfer to the production federation in 2005/6. This would populate the federation with enough members to be the main environment for use by institutions of all kinds.

Legal issues

Only IPR (Copyright) and Data Protection should affect the setting up of a simple Shibboleth connection.

Data Protection

Institutions setting up a Shibboleth Identity Provider server must be aware of the implications of Data Protection Law.

The law governing the way in which personal data should be used of processed is set out in the Data Protection Act 1998, which came into effect on 1 March 2000. Personal data can be collected from individuals and processed provided there is compliance with the eight data protection principles[18]. When institutions collect data from users (staff and students) they need to be clear, and unambiguous, about the ways in which that data will be used, including its passing to third parties involved in authentication or authorisation.

It is worth noting that in the case of a Shibboleth connection, no sensitive personal data which is covered by the act should be required to authenticate and authorise users.

Intellectual Property Rights

[18]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

Users must stay within the law of copyright when using electronic journals.

E-journals are subject to copyright law in exactly the same way as printed material. E-journals also tend to have additional restrictions on use defined by their associated license agreements. The main prohibited activities include:

- downloading entire volumes or issues of journals;
- the commercial use of journal content.

It is unclear exactly how developments in the management of rights in the digital environment will alter the availability and use of digital works. The increasing enclosure of digital works by publishers, supported by changes in the legal environment that favour rights-holders over public access, has seen innovative responses from those who wish to maintain the free flow of information, in the form of creative copyright licensing (open source licensing, the Creative Commons model), new publication models (pre-print publishing e.g. SSRN, open access publishing) and greater co-operation and collaboration between interest groups, such as educational institutions.

For more detail on Digital Rights Management and legal issues surrounding Shibboleth Federations see *An introduction to Shibboleth Federations* at http://iamsect.ncl.ac.uk/deliverables/.

Shibboleth and e-journal access: case studies

There is only one UK implementation of shibbolised access to e-journals which has been written up (Section 4.1). However, the JISC funded Core Middleware[19] Development programme (of which this project is a part) will investigate uses of Shibboleth and explore ways of extending the Shibboleth architecture, for example to incorporate more sophisticated digital rights management capabilities.

Athens, which was designed for use with digital libraries, has been successfully integrated with local institution authentication schemes (this variant of Athens is described as AthensDA, or Athens with Devolved Authentication) and allows integration of an institution's learning environment with that institution's external electronic subscriptions. The JISC believes that the same degree of integration, or better, will be possible with Shibboleth but a number of the key concepts, such as the Internet2-MACE CourseID[20] parameter definition, have not to date been tested in production use. In 2005, Athens will extend the functionality of its AthensDA software to incorporate the Shibboleth architecture and in particular the SAML protocol for attributes. This work will be developed under the aegis of the JISC Core Middleware Infrastructure programme.

SECURe

The LSE Library has over 3,000 e-journal subscriptions. At the moment, the licensing and access information attached to LSE electronic resources is somewhat dispersed. In part due to its size and number of users, the LSE library has proved to be a difficult operational testbed. Changes to instructions to users, such as how to login, are only changed once a year, so development projects such as SECURE have to fit with this schedule.

The SECURe project implemented Shibboleth for cross-domain access management; initially (from January 2004) LSE users accessed e-journals in the Jstor repository via a Shibbolised channel, instead of Athens mediated access to a UK mirror of the repository. Jstor has been one of the service providers participating most actively in Shibboleth development, amongst their motives being a reluctance to continue the costly practise of maintaining mirror repositories in the UK and other places, and a recognition that the widespread use of insecure IP proxies by universities (to enable access for off-campus users)

[19] http://www.jisc.ac.uk/index.cfm?name=programme_middleware
[20] http://middleware.internet2.edu/courseID/docs/draft-internet2-mace-courseid-scenarios-00.html

was responsible for significant 'leakage' of resources to unlicensed users. MIMAS, a JISC data centre, hosts the Jstor mirror in the UK. Although the Shibboleth connection has been operational, there have not been a lot of users for the service.

NSDL

The National Science Digital Library is using Shibboleth for authentication and authorization to resources available via the NSDL portal that are restricted to certain audiences.

Users register and login to the NSDL through the Access Management System developed by Core Integration partners at Columbia University. The system enables federated identity management allowing web content providers to establish relationships with subscribers on an individual, institutional, or other basis. Subscribers can locally manage their personal or institutional data.

In the first release of the NSDL, content providers (origin sites) and subscribers (target sites) must have a pre-existing peer relationship. To develop a Library with the depth and breadth of resources required by teachers and students the NSDL will continue to develop an Access Management System that provides benefits for both users and information providers. A growing variety of content that is both open to users and protected by content providers will create a Library that meets the science, technology, engineering and mathematics educational needs of students and teachers.

SwitchAAI

The Swiss team do not to date (June 2005) have success stories to report regarding access to e-journals using Shibboleth. That is mainly for two reasons:

- It has taken SwitchAAI much longer to get the federation member service agreement finalised than anticipated. That was a pre-requisite before putting the federation partner agreement together which is a pre-requisite for adding external partners to the Federation like e-publishers. This should be finalised by end June 2005 so that SwitchAAI can contact publishers.
- The other route followed by SwitchAAI is with EZproxy[21], which is now shibbolized, but that activity only started in June 2005. The server has been set up and will be configured by July 2005.

InCommon

In the US, within the InCommon Federation, ScienceDirect (which offers Elsevier content) is operational. Initially, the implementation on ScienceDirect was piloted with selected universities from the InCommon Federation. From December 2004 utilisation of Shibboleth software has meant that users from all InCommon universities and colleges can now benefit from remote off-campus access to ScienceDirect without additional administration by participating institutions.

[21] http://www.usefulutilities.com/