From a pile of IP addresses to a clear authentication and authorization with Shibboleth

BERND OBERKNAPP
ReDI Development & Operations Manager
University of Freiburg

ATO RUPPERT
Library IT Director
University of Freiburg

FRANCK BOREL
Software Engineer
Common Library Network
Göttingen

JOCHEN LIENHARD
Software Engineer
University of Freiburg

The authors, who all work in Germany, are from left to right: Bernd Oberknapp, Ato Ruppert, Franck Borel and Jochen Lienhard

It is common practice today to use IP access control. This seems to be an easy-to-use method, but in fact there are a lot of questionable issues connected to it. A Shibboleth-based authentication and authorization infrastructure (AAI) offers a solution for these issues. For this reason the AAR project was begun in 2005 with the goal of building a Shibboleth infrastructure for both German higher education and research and for the Internet portal for scientific and scholarly information called vascoda. Within just two years, the project attracted a very high level of interest. This article describes how we built the Shibboleth infrastructure and it will show the current status of the DFN-AAI federation for German higher education and research.

Limitations of IP access control

Today, access to licensed resources is most often controlled through IP access control. At first sight this may look like an easy and convenient method, but on taking a closer look it turns out to be a rather insecure method, severely limiting the options for authorizing access to resources, especially when private networks, proxies and virtual private nets (VPNs) are involved, requiring rather special solutions for off-campus access. Proxies and VPN gateways are often shared by departments of a university, making it impossible to license a resource with IP access control for just one department without excluding proxy and VPN access, in many cases excluding remote access and even wireless networks at the same time. Furthermore, the network structure at most higher education institutions (at least in Germany) has been built without IP access control in mind. This often results in long lists of IP ranges that have to be kept current, resulting in a nightmare if you have to do this for dozens or sometimes even hundreds of institutions or publishers. But the worst problem may be seen with remote access, especially in countries like Germany, where students do not usually reside on the campus. With cheap broadband network access available, and more and more faculty and staff also working off campus, the remote access problem has become an increasingly urgent issue to tackle in recent years.

All methods used for remote access via IP access control like VPNs, proxies, rewriting proxies and proprietary methods offered by some publishers have severe disadvantages, with most of them failing in critical situations. Proxies and VPNs require a special configuration on the client side (the user's computer) resulting in lots of technical problems, especially in connection with VPNs. If the client is located behind a firewall, the institutions' proxies and VPNs cannot normally be used. Rewriting proxies do not suffer from these problems, but for them, special URLs are required in order to be able to access the resources. Maintaining thousands of special URLs requires a lot of additional organizational effort from the library, and this method will not work for links the library cannot control, such as direct reference links from one publisher to another.

vascoda and the AAR project

vascoda[1] is an Internet portal for scientific and scholarly information, offering interdisciplinary and subject-specific search. More than 40 academic libraries, particularly those with specialized collections, central subject-specific libraries as well as information institutions of national importance in Germany, participate in vascoda. The data providers and partners deliver their content materials via virtual subject-specific libraries, information associations and other subject-oriented portals.

A flexible, distributed solution for authentication, authorization and access control is essential in a complex environment like vascoda. Therefore, the Federal Ministry of Education and Research has funded the AAR infrastructure project[2] for vascoda. AAR stands for authentication, authorization and rights management (in the sense of licence information management, not DRM). The project was executed by the University Library Freiburg, which runs the ReDI database service[3] (see below), with an emphasis on authentication and authorization, and the University Library Regensburg, which hosts the Electronic Journals Library[4] (EZB), emphasizing licence management and authorization. In this article we present the results from the Freiburg part of the project, which started in January 2005 and ended in June 2008.

Shibboleth[5] was the technical solution of choice for us, not only because it is a standards-based, open source, inter-institutional web single sign-on (SSO) solution that can solve the remote access problem, but also because it was already gaining more and more support both from higher education institutions in the US and Europe and from important publishers and aggregators such as EBSCO, Elsevier and JSTOR. In addition to the SSO functionality, Shibboleth provides an attribute exchange framework with a strong focus on privacy protection. The identity provider, run by the user's home institution, uses the institution's identity management system so that attributes (e.g. the user's relationship with the institution) determine the user's access rights to the services. The service provider receives a set of attributes from the identity provider, agreed upon by the institution and the service provider, allowing him to make informed authorization decisions for individual access to the resources. The identity provider cannot only release different sets of attributes (depending on the service provider) but it also allows the users to control the attribute release with a tool like SWITCH's uApprove[6].

The original AAR project plan included a) building a Shibboleth test environment, b) 'shib-enabling' ReDI and the vascoda portal, c) convincing higher education institutions in Germany and (commercial) service providers to implement Shibboleth and supporting them in doing so, and finally d) proposing a long-term solution for using Shibboleth in German higher education and research. In mid-2005 it became clear that we needed a federation for German higher education and research for working with international services providers as soon as possible. Fortunately the Deutsches Forschungsnetz (DFN), Germany's National Research and Education Network, agreed to create the DFN-AAI federation[7] in co-operation with us. Through this the AAR project became the testbed for the DFN-AAI.

Getting started with Shibboleth

After familiarizing ourselves with the Shibboleth software, we set up the AAR test environment with multiple identity and service providers. During this phase we had some very helpful discussions with our colleagues from SWITCH, who already had a working federation based on Shibboleth at that time. In October 2005 we started with a series of AAR workshops (two per year) to promote

Shibboleth in Germany and help institutions and service providers to get started with Shibboleth.

At the end of 2005 we were ready to go into production with some internal applications like the monitoring of the library servers and services (Nagios), our internal web-based backup service (BackupPC) and selected library applications. This allowed us to demonstrate the advantages of Shibboleth in a local production environment, especially single sign-on, which allows users to access all services with a single account and just one login, and the access control based on user attributes. This helped in convincing the University Computing Center, the University Hospital Computing Center and the university administration to establish a central SSO service for the university, dubbed 'myLogin'.

ReDI – the pilot federated application

For us, the next major step was to shib-enable ReDI (Regional Database Information Services), the central database service for the higher education institutions in the federal state of Baden-Württemberg. ReDI is funded by the Ministry of Science, Research and the Arts and has been in service since 1999. Today, ReDI offers access to almost 700 databases including about 400 Windows-based databases (on Windows terminal servers under the Citrix Presentation Server for web-based access), including CrossFire Beilstein and Gmelin hosted on ReDI server systems in Freiburg and Stuttgart. ReDI is used by more than 50 sites in Baden-Württemberg and an increasing number of sites from other federal states in Germany.

In 2005 ReDI already had a proprietary authentication and authorization system based on a proprietary protocol that was able to authenticate users against local user databases in their institutions and to authorize access to databases based on user groups. So we had the choice either to add Shibboleth as an additional authentication and authorization method or to completely replace the old proprietary system with Shibboleth. We decided to go for the latter and, for a quick start, to set up Shibboleth identity providers on the ReDI servers for all sites offering username and password access to ReDI, with the exception of the University of Heidelberg, which already had its own identity provider. These identity providers still used the old proprietary protocol for accessing the local user databases. This approach allowed us to go into production with the shib-enabled ReDI as early as April 2006.

Even though we were using Shibboleth at the federal state level at that time, we decided not to build a formal ReDI or Baden-Württemberg federation, because the DFN had already agreed to build the DFN-AAI federation for German higher education and research institutions together with us.

vascoda – personalization with Shibboleth

The vascoda portal is based on the IPS portal software run by the Hochschulbibliothekszentrum North Rhine-Westphalia (HBZ) in Cologne. When we started building a search portal for the University Library Freiburg at the end of 2005, we decided to also use the IPS software and to shib-enable IPS in co-operation with the HBZ. During 2006 we added not only Shibboleth authentication and authorization to IPS, but also personalization with Shibboleth: every time a user logs in to the portal, the identity provider sends a unique pseudonym for the user (a persistent name identifier) that allows the portal to recognize the user without revealing the user's identity. This way, users do not have to register with the portal for using personalized features, resulting in an even smaller number of accounts for the user and no hassle with lost accounts or passwords for the service provider.

The University Library Freiburg portal went into production in February 2007. The vascoda portal has been using Shibboleth for personalization since the end of 2007; the work on authorization for access to licensed content in the vascoda portal is still in progress.

Building the DFN-AAI federation

Building a federation is a complex process that not only involves implementing technical systems including metadata administration, test environment and a central discovery service, but it also involves setting up a legal framework, including policies and a service provider contract. This contract covers the use of the federation infrastructure, not the licensing of content. Existing federations like SWITCH AAI in Switzerland and HAKA in Finland were good examples of what the requirements

were. The people from SWITCH were once more very helpful, allowing us to use an adapted version of their service provider contract for the DFN-AAI. Nevertheless, this process took almost two years. The pilot phase started in May 2007, and the DFN-AAI has been in production since November 2007.

Current status and future developments

The implementation of the myLogin SSO service for the University of Freiburg has initiated some improvements in the University's identity management. This is an ongoing development: while we are adding more and more applications from different areas (library, e-learning, administration, etc.) further improvements will be necessary, especially for the user groups and privilege management. The upgrade to Shibboleth 2, which is planned for the forthcoming months, will allow us to add applications that require features not available in Shibboleth 1.3, including reauthentication, for example, and it will improve the interoperability with other implementations based on OASIS' Security Assertion Markup Language (SAML), such as simpleSAMLphp[8]. For the library services, the long-term goal is to get rid of that pile of IP addresses and completely replace IP access control with Shibboleth – not only for remote access but also for on-campus access.

In Baden-Württemberg, so far only the Universities of Freiburg, Heidelberg, Konstanz and Tübingen are running their own identity providers and are members of the DFN-AAI. We are therefore still running a lot of the 'quick start identity providers' that were set up in 2006 on the ReDI servers. However, this is a temporary solution, and we announced last year that this solution will only be supported until the end of 2009. This sparked off a lot of activity in the higher education institutions in Baden-Württemberg. Some more universities are now ready to join the DFN-AAI. At the teachers training colleges a project has started to improve the identity management, enabling some of them to join the DFN-AAI until the end of the year. Some universities of applied sciences will probably join the DFN-AAI in 2009, too. The main issue for most higher education institutions, not only in Baden-Württemberg, but also in the other federal states in Germany, is identity management. There are lots of projects for improving identity management which should make many institutions ready for joining the DFN-AAI, but this process needs some more time. Nevertheless the number of identity providers in the DFN-AAI is growing steadily.

Some important publishers have already joined the DFN-AAI, but a lot of publishers are still missing from the federation. Fortunately, the number of publishers that support Shibboleth is growing rapidly, most notably because of the transition from Athens to Shibboleth in the UK. We are continuously working on adding more publishers to the DFN-AAI, and even though the AAR project ended in 2008, we will continue to support some publishers in implementing Shibboleth.

From the user's perspective the single sign-on for many applications and off-campus access to licensed resources without proxies and VPNs is a remarkable improvement. Unfortunately, Shibboleth introduces an inconvenience for the user, the "where are you from" (WAYF) or discovery service (DS): service providers do not usually know the user's home institution and therefore use a WAYF to acquire this information from the user. For service providers with users from multiple federations, currently the only practical approach is to build their own WAYF, so the user has to select the home institution over and over again in different WAYFs. It is usually possible to skip a WAYF with a special 'WAYFless URL' that includes the home institution information, but unfortunately these WAYFless URLs are complex and not very stable. Therefore, we are working on an upgrade for ReDI that will allow us to use WAYFless URLs internally and still provide stable and less complex URLs that the libraries can use to link the resources.

Conclusion

Building an AAI is a long-term process. The AAR project was very successful in triggering that process for German higher education and research, but the process is far from over and the adoption has just started. Shibboleth was the right choice, because it is rapidly gaining more and more support in higher education institutions, from publishers, and from application developers worldwide.

Apart from shib-enabling more applications the main point we have to work on in the future is in

improving user experience. How do we deal with the (multi-federation) WAYF problem? And how can we deal with (single) logout? Another important issue we have to work on is usage statistics for licensed content. Today, usage statistics are often broken down by IP ranges for accounting. How can we handle this with Shibboleth? An accounting attribute could be a solution.

We are looking forward to continuing to work with an expanding Shibboleth community.

References

1. vascoda portal: http://www.vascoda.de/ (Accessed 27 January 2009)
2. AAR project homepage (German only): http://aar.vascoda.de/
3. ReDI database portal (German only): http://www.redi-bw.de/ (Accessed 27 January 2009)
4. Electronic Journals Library (EZB): http://ezb.uni-regensburg.de/ (Accessed 27 January 2009)
5. Shibboleth project homepage: http://shibboleth.internet2.edu/ (Accessed 27 January 2009)
6. SWITCH AAI tools: http://www.switch.ch/aai/support/tools/ (Accessed 27 January 2009)
7. DFN-AAI homepage (German only): http://www.aai.dfn.de/ (Accessed 27 January 2009)
8. simpleSAMLphp homepage: http://rnd.feide.no/simplesamlphp (Accessed 27 January 2009)

■ **Bernd Oberknapp**
ReDI Development & Operations Manager
University Library Freiburg
E-mail: bo@ub.uni-freiburg.de

■ **Dr Jochen Lienhard**
Software Engineer
University Library Freiburg
E-mail: lienhard@ub.uni-freiburg.de

■ **Ato Ruppert (AAR Project Leader)**
Library IT Director
University Library Freiburg
PO Box 1629
79016 Freiburg, Germany
Tel: +49(0)761 203 3906
Fax: +49 (0)761 203 3987
E-mail: ruppert@ub.uni-freiburg.de

■ **Franck Borel**
Software Engineer
Head Office of the Common Library Network
Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen, Germany
E-mail: borel@gbv.de