

METASEARCH

Issues with Authorization and Authentication

Abstract

Database providers generally require that users identify themselves as members of a population that is licensed to use that database. Providers do this in order to restrict access only to paying customers, in order to maximize the value of their investment in creating the database. A number of mechanisms exist for user authentication, ranging from “traditional” approaches such as IP address validation to newer technologies such as Shibboleth. Even after a user is authenticated as part of a group entitled to access an electronic resource, authorization of that library’s right to access the database under the terms of the license must be handled.

Introduction

Library users are confronted with far more information supply choices now than ever before. Libraries subscribe to an increasing number of citation and full text databases, spending hundreds of thousands of dollars annually to provide electronic information to their service populations.

Library users, while eager to have access to these new sources of information, may be resistant to learning the specific skills necessary to successfully search each database source. Further, library users have neither the patience nor the inclination to search repeatedly in different databases to satisfy their information needs.

Many library users – particularly those in public libraries - want a service where they can enter a search argument a single time, and cause the underlying search apparatus to search multiple electronic data sources. In addition, they want the results from that search to be collected, organized, and displayed in a way that the most relevant retrieval is presented first. Finally, they want a simple and integrated way to locate, retrieve, display, and/or print the full text of the resources that they have located. In short, the library public wants a “one-stop-shopping” interface to the entire range of library accessible databases.

Metasearch services have evolved as a solution to this need for one-stop information retrieval.

The Problem

Thousands of databases are available to libraries to license, created or distributed by hundreds of different publishers. Since the databases represent valuable and saleable intellectual property, database publishers have a justifiable

interest in restricting access only to those who have paid the contractual license fee for access.

The traditional library mission is to deliver information to the end user. In previous years, before the "digital revolution", the physical library was often the central delivery point for information delivery. With the evolution of the Internet and the concomitant increase in digital literacy among many library users, the physical library location has become irrelevant. Library users expect that their libraries will contract for a wide variety of electronic resources, and that these resources will be available to them without respect to the specific physical location of the library user.

Current state: Access management

One of the current challenges of Access Management in the library user space is the variety of methods used to qualify users to access protected resources. For this discussion, it is helpful to define each of the steps used in this qualification process. We use "authentication" to signify the act of identifying a user in a trusted database, such as the patron table in an integrated library system (ILS) or the campus student database. (This process may also include "validation", often used to ensure the user is a valid recipient of the resources subscribed by the library, but for purposes of this discussion, we will treat validation as the same step as authentication.) Once the user has been authenticated, the last step is to "certify" to the database vendor that the user falls under the guidelines of the subscription agreement.

Methods of Authentication

Although authentication standards, like NCIP in the library space and LDAP for directory services, are emerging, we know of no integrated library systems that currently use the standards for this purpose. Those who provide authentication do so using proprietary methods, using their own tools to authenticate the user against the appropriate patron database. A few individual libraries use LDAP to authenticate users, but this standard has yet to be widely implemented in turn-key systems.

Current Methods of Certification

In our brief survey of database providers, we discovered that, again, no standard is used to certify users "worthy" of database use. However, we found that all vendors who accommodate remote users use one of four methods and we describe each below.

IP-Filtering

Virtually all database providers rely, to some extent, on IP recognition or "filtering" as the primary, and sometimes only, method of proving the user is valid. This means the user must either be physically in the library or must be accessing the database through a library-configured proxy server. While there are challenges in using IP-filtering, such as possible IP-spoofing and the perpetual hassle of keeping the IP lists synchronized between the library and the vendor, this is still the most widely used method of certifying that the user is a "member" of the library and has rights to use the vendor's database.

Referring URL

When an authenticated user selects a protected resource, the URL from the page from which he is launched or "referring URL" is passed to the database provider in the HTTP header. If a service is provided which allows only authenticated patrons to access the referring URL page, the database provider can be assured that the user has been authenticated before accessing its services. If this method of certification is used, the database provider (or protected resource) must maintain and recognize referring URLs for each of its customer libraries. This method is somewhat complex, requiring the library to discover and report all valid referring URLs to the vendor, and requiring the vendor to record and recognize valid Referring URLs for each library. We found that most database vendors who conscientiously provide access to remote users (above and beyond IP-filtering) accommodate this method.

URL-Embedded Username and Password

A method preferred by other database providers is one where it assigns the library a username and password, which are placed as variables in the "Success URL" or the URL which is used to access the database, once the user has authenticated. Since, once again, access to this link is restricted to those who have successfully authenticated, the database provider can be relatively assured that the user is a qualified patron of a subscribing library. The disadvantage of this method is that if care is not taken, once a user has authenticated and gained access to the vendor's page, he could possibly save the password-embedded URL as a bookmark and subsequently use or misuse it.

Database Vendor provided Script

A few database providers provide scripts which encrypt or otherwise securely communicate certification information in the HTTP message. Some libraries and/or library systems execute such scripts after the user has successfully authenticated, providing access to the desired database.

Additional Vendor requirements

Some database vendors, to further restrict access to their databases, require the entry of a library card number (or other unique identifier) for access. This additional step is generally not validated against the library's user database. Instead, the structure of the unique number is examined for length and whether that number begins with a prefix appropriate for that library.

Current State: Authorization Management

Even if a specific user has identified himself to the library and the remote electronic resource as a "valid" member of the population contractually entitled to access that resource, other factors may come into play that may limit access. Many electronic resources are licensed on a "per simultaneous use" basis, in which a library pays for an estimated maximum number of uses. This is usually based on an algorithm that takes into account the potential user population, the number of physical locations from which access may be made, and the potential user audience for that resource.

If a library has contracted for ten simultaneous uses of database X, the first ten users will be accommodated and served without a problem. However, when the eleventh person in the library system attempts to access database X, controls must be invoked that restrict his access.

There appears to be no standard approach to this type of authorization control. Generally, the remote electronic resource tracks simultaneous usage within a given library and returns a "unable to authorize" message to the local library. However, some more sophisticated local authorization and control systems try to manage simultaneous users locally, and only permit the "licensed" number of sessions to be completed to the remote electronic resource.

Any metasearch system needs to be able to manage the authentication process and send appropriate messaging back to the requesting workstation when the macro-authentication process fails.

Potential Solutions and Approaches to Issues of Authentication

- Proxy Servers
- Individual Passwords per user
- Individual Passwords per institution
- IP address validation
- Referring URL page validation
- Embedded passwords in HTTP message
- "campus-wide" authorization services such as CAS at Yale and WebAuth at Duke

- X.509 certificate exchange on an individual basis
- PAPI
- TEQUILA
- Shibboleth

Further reading: (selected list, by no means exhaustive)

Federated Digital Rights Management :

A Proposed DRM Solution for Research and Education

<http://www.dlib.org/dlib/july02/martin/07martin.html>

(Article in D-Lib Magazine discussing approaches for DRM in supporting educational goals)

Authentication to Licensed Resources (JSTOR)

<http://uk.jstor.org/about/authentication.html>

(discusses JSTOR's approaches to authentication)

Access Management for Networked Information Resources by Clifford Lynch

<http://www.educause.edu/ir/library/html/cem9842.html>

(overview article)

Authorization/Authentication for Patron Remote Access to Electronic Resources

(powerpoint by Kerry Bouchard)

<http://libnt2.lib.tcu.edu/staff/bouchard/ugc2000/remotearchive/sld001.htm>

(useful visual introduction to issues relating to authorization)

A White Paper on Authentication and Access Management Issues in

Cross-organizational Use of Networked Information Resources by Clifford Lynch,

editor (cliff@cni.org)

<http://www.cni.org/projects/authentication/authentication-wp.html>

Digital Library Architectures, Systems, and Tools

<http://www.diglib.org/architectures.htm>