Remote Patron Validation: Posting a Proxy Server at the Digital Doorway

"... just like a good hotel or apartment building, we need security in place to protect our users—and increasing use of remote access is making patron authentication more important than ever."

by Peter Webster

Delivering online services outside of library buildings and off-campus at universities and colleges is quickly becoming a major and essential part of what librarians do. Remote access to e-journals and other online services from users' homes and offices is very popular and is growing in demand. In academic libraries it is common to offer campuswide and remote access to thousands of online journals, online reference works like encyclopedias and directories, and, increasingly, electronic books as well. Public libraries are also offering remote access to a growing selection of e-journals and other online resources.

Librarians and online vendors have come a long way with remote access in a very short time. To do so, we have developed numerous complex hardware and software infrastructures. The methods we have used have been haphazard and to some extent accidental, and they have not made the best use of our integrated library systems. Moreover, just like a good hotel or apartment building, we need security in place to protect our users—and increasing use of remote access is making patron authentication more important than ever. So far, this has involved setting up special proxy servers, maintaining multiple password lists, or arranging remote access one online product at a time. We have sometimes even required each user to modify his home or office Web browser in order to access library resources.

Better methods of offering remote access are developing, making the process easier for users, more reliable, and less complex and costly for libraries. With help from both integrated library system (ILS) vendors and online product vendors, we can continue to standardize and improve remote access to truly provide a "library without walls."

In this overview I will take a look at library remote access and the methods we currently use for it. In particular, I will look at the importance of proxy servers. I will outline the different kinds of proxy servers as well as the various alternatives to them. I'll look at developments in proxy serving and consider some improvements that are still needed, particularly user validation from third-party servers and better direct-validation methods from online vendors. Finally, I will discuss where we might expect remote access to go in the near future, and how we can better guard our vulnerable doorways.

For the last 10 years, I have looked after IT for the Patrick Power Library at Saint Mary's University, a medium-sized university in Halifax, on the east coast of Canada. (A lovely place to visit in the summer, by the way.) I think our experience with online resources has been similar to that of most academic libraries. We started offering online journal indexing via CD-ROM in 1991. Campuswide LAN network access came in 1993; by 1996 we offered slow and unreliable off-campus access to a few databases via telephone modem. As the World Wide Web developed, we took advantage of the full-text resources becoming available over the Internet, and the range of online resources we offered grew (and continues to grow) rapidly. With Internet accessibility came the obvious possibility of offering remote access via the Internet to our users.

Up until 1999, one librarian handled all database developments part-time at Saint Mary's, with occasional support from hard-pressed university IT services. The addition of a full-time library systems technician in 1999 was a critical step in the development of our online services, and made it possible to maintain a dedicated proxy server. Though the cost of online databases is high, the cost of our proxy services has been relatively small. Our first proxy server was developed in a matter of months, using old computer parts and free software. We have added better hardware and additional software as funding became available. But our current investment for two Dell servers and software is less than $8,000.

> "The proxy server is not unlike the modern librarian, serving as a helpful and discreet intermediary between users and online information. But it also has an important role as a doorman, ensuring that only the right people are allowed to pass."

User Validation Is Critical

Several key components have been necessary to make large-scale remote access possible. First, of course, was the Web becoming widely available in the homes and offices of our library users. The second key component was IP (Internet Protocol) number validation for accessing online databases. Validation, or what is also called user authentication, is critical to the use of online resources. When vendors license a library to use their valuable online products, they must ensure that only that library's users can access those products. Without reliable validation, costly products would soon become freely available on the Web.

Assigning individual user names and passwords is a standard method of securing online database and servers. But issuing individual user names and passwords for each library database is impractical for libraries on-site, much less for remote access. Our relatively small university library offers more than 40 databases, and hundreds more individual e-journal titles. We could not begin to manage so many passwords or ask our users to tolerate them.

The Importance of IP-Based User Validation

The key alternative to password control has been IP validation. It continues to be the most practical method for securing and validating access to the online products libraries offer. It has become the standardized method for large-scale user validation. *CIL* readers will know that every computer using the Internet is assigned a unique 12-digit IP number. All computers in an organization, such as a university or library, will often have a common sequence of IP numbers. For example, all PCs and servers on our campus have IP numbers with the same first six digits. Internet servers have the ability to restrict access only to computers with particular IP numbers or ranges of numbers. So vendors can set up their database servers to allow access from any computer with the IP number of our university campus. Anyone on Saint Mary's campus can then access any online product we pay for without the need to log in with a password.

In the last couple of years libraries have struggled with vendors to get them to offer IP validation as a standardized alternative to password control—and for the most part, the vendors have begun doing so. However, a few small database providers and many e-journal publishers still offer only password access, and some e-journal publishers charge extra for IP-controlled access. IP validation was not developed as a means of authenticating database use; it has been adapted for that use largely as a fortunate accident.

How Proxy Servers Help

With IP validation becoming the preferred method of securing and validating database access on-site in libraries, people needed a method to do the same for remote library users. The method that has become standard is the proxy server. A proxy server is a computer on-site at a library that users can connect to over the Internet. This server acts as an intermediary between the remote users and the database servers that the library makes available. The remote users cannot access the vendors' databases directly from their home PCs. But they can connect to the proxy server, which then passes information back and forth from the remote users to the vendor database, making it appear as if they are working from valid IP numbers at the library rather than from their homes or offices. This "fooling" of online vendor servers is, of course, done with the full knowledge of the database vendors. Online license agreements now commonly authorize the use of proxy servers. The proxy server is not unlike the modern librarian, serving as a helpful and discreet intermediary between users and online information. But it also has an important role as a doorman, ensuring that only the right people are allowed to pass. So proxying requires a reliable method for validating the users who connect.

A recent survey by Peter Murray, computer services librarian at the University of Connecticut School of Law, is one of the first to explore the use of proxy servers in libraries.[1] Though it is a very small survey, it shows the prevalence of proxy servers in libraries, and it provides valuable information about how they are being used, both for remote access (far and away their most popular use) as well as for other purposes.

Because of the costs and skills required to operate a proxy server, we have turned to them somewhat reluctantly. But for now, proxy servers of different kinds are being used very successfully to extend online services beyond our walls. In my library, we have grown used to occasionally serving online users in Nunavut, in Canada's far north, or in Ireland, or in Gambia, Africa.

As proxy servers have developed, a number of different approaches have been used that are worth taking a look at:

Browser-Directed Proxies (Diagram 1): A common method of proxy service requires each remote user to change the setup on his or her Web browser to point to the address of the proxy server. The change is only a matter of inserting a proxy server's URL in the appropriate location in the browser's preferences. But the steps required to accomplish this vary depending on the browser software vendor and version differences. Libraries must provide setup instruction for many different versions of Netscape, Internet Explorer, and other browsers. The open source software Squid is one popular example of browser-directed proxy software. It is a very reliable and versatile proxying solution, and was the most common kind of proxy server until recently. But the extra user support it needs in order to make changes to users' home Web browsers has been one of the key problems with this method. At Saint Mary's we explored using Squid, but since we have very limited (though brilliant) systems support, we wanted to avoid tinkering with users' Web browsers.

Server Software Proxies (Diagram 2): An alternative to browser-directed proxies are server software proxies, which don't require any alteration to the user's browser. These are simply software applications, often written in the Perl server scripting language, which run on a Web server to act as intermediary between remote users and online resource vendor servers.

A number of libraries have programmed their own proxy serving software. This is the route my library took a few years ago when we began making databases available to our users over the Internet. There is software available on the Internet intended to pass network traffic through an intermediary server, so the location and Internet identification of that traffic appears to be from the intermediary server rather than the true source. There are freely available Perl scripts that do this, both for legitimate reasons of privacy and for less-legitimate purposes. The Perl programming wizard on our staff was able to alter existing freely available Perl to allow our Web server to proxy Internet traffic between users and vendor databases. We set up a validation script so that only legitimate users to whom we gave passwords could log in to the proxy. This script merely contacted a list of database Web sites and passed their output strings unaltered to the remote user's computer, to be processed and displayed by the user's Web browser. It took only a few days to have a basic working application. I have to confess that we know very little about the details of what a database server might pass to remote client Web browsers. The homegrown software had its share of bugs, but it worked well for Saint Mary's for several years.

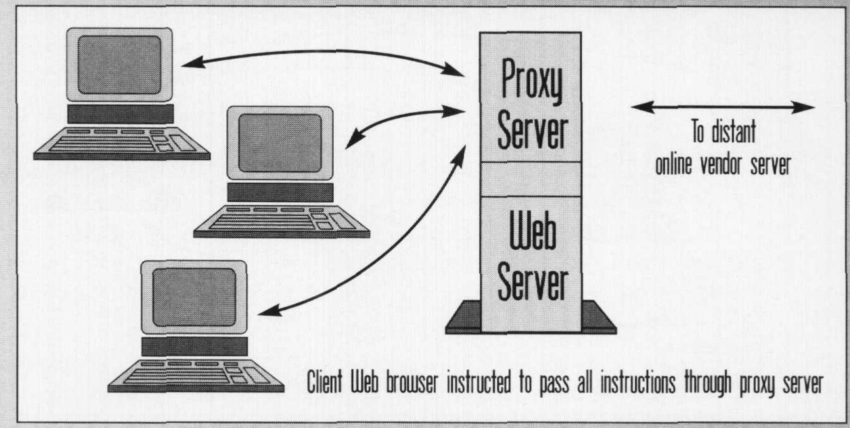
Proxy Server
Web Server
To distant online vendor server
Client Web browser instructed to pass all instructions through proxy server

Diagram 1: A browser-directed proxy server

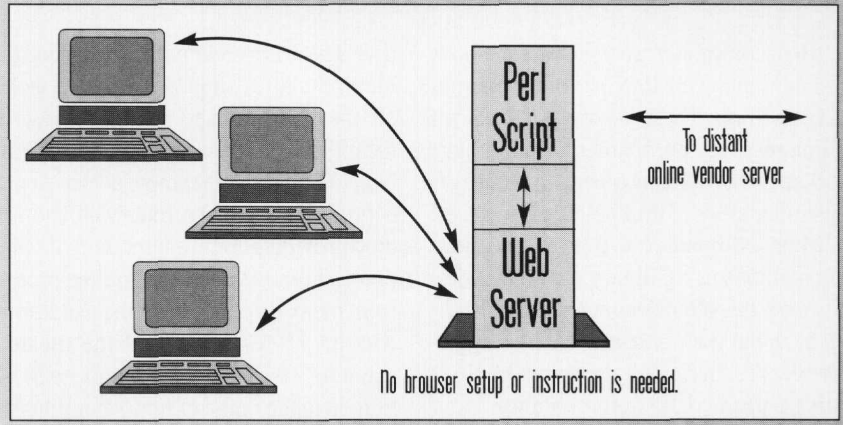
Perl Script
Web Server
To distant online vendor server
No browser setup or instruction is needed.

Diagram 2: Server software proxy

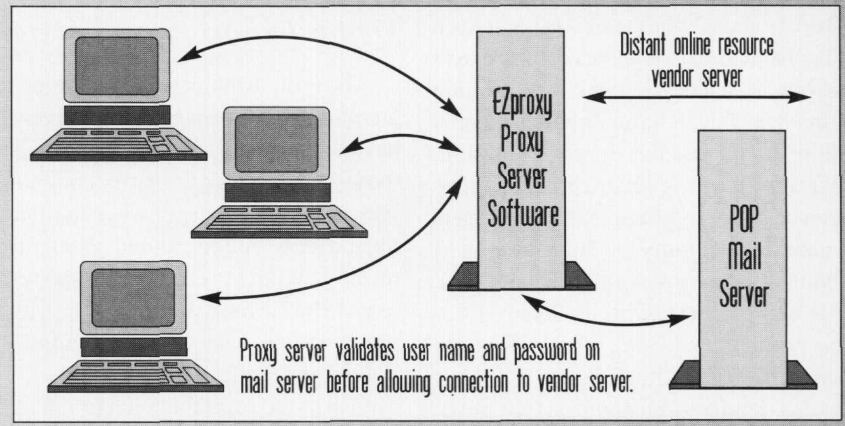
EZproxy Proxy Server Software
Distant online resource vendor server
POP Mail Server
Proxy server validates user name and password on mail server before allowing connection to vendor server.

Diagram 3: EZproxy proxy software using third-party server validation

EZproxy emerges as a standard in proxy software (Diagram 3): In 1999 we changed over to commercial software, Useful Utilities Co.'s EZproxy. It does what our homemade proxy software did, but it was written by someone with greater expertise than we had, and who had the time to develop and customize the software as particular problems were encountered and as new online products came along. This software has rapidly become the most popular proxy method in libraries. It has replaced browser-directed and custom-programmed proxy software in many libraries. EZproxy was the most popular proxy software reported in Peter Murray's survey. There are several other commercial proxy products. Innovative Interfaces' Web Access Management software is popular, and Obvia Corp. offers its Remote Database Access product. But at this time EZproxy is setting the standard. Nearly half of survey respondents were using it for remote access. It was more than twice as popular as the nearest competing proxy product.

Many Vendor Alternatives

Proxy servers are not the only means of providing library users with remote access. Many database vendors offer methods for validating remote users as an alternative to running a proxy server, for which many libraries lack the necessary resources. I have worked with both public and academic libraries where even the funds needed for a small proxy server are not available. More importantly, they didn't have the IT knowledge and personnel they needed to maintain a server. For many libraries in this position, direct vendor methods of validation are the only remote-access options available. Database companies such as ProQuest, EBSCO, and Gale Group offer a range of options for remote user validation.

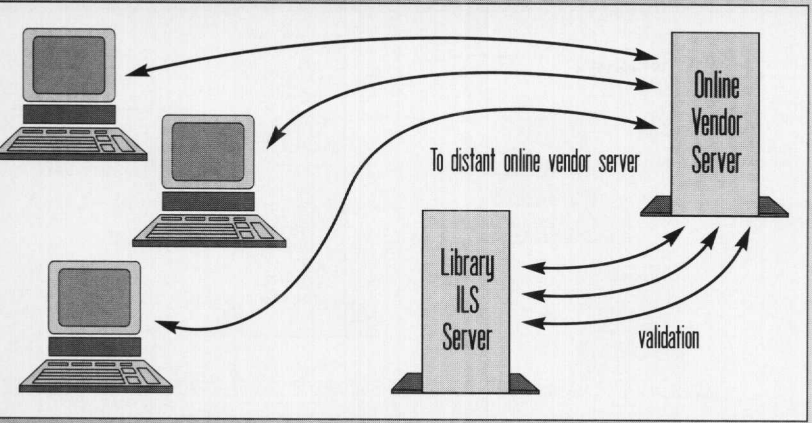

Diagram 4: Direct third-party validation by an online vendor server, using existing patron information

Most commonly, they provide methods for maintaining custom lists of user names and passwords. These are some of the larger online resource companies, which I have worked with, but many other companies offer similar kinds of remote-access services.

Many companies also provide validation methods using library card bar codes. However, this often does not involve checking each bar code against live library patron records. Instead, it uses what EBSCO calls "patterned ID authentication" and what Gale's InfoTrac calls "barcode pattern analysis." This method works as follows: When a remote library user is asked to type in her bar code to log in to a database over the Web, it is checked to see that a portion of the bar code, a number pattern that is unique to a particular library, is present. This is not the most secure practice, since it can be quite easy for unauthorized users to make up a dummy bar code, as long as it contains the correct unique digits.

Both ProQuest, with its Secured Access product, and Gale, with its Remote Patron Authentication Service, use a system that develops a customized CGI proxy authentication script for each library. Libraries then load this CGI software onto their own Web servers, where it provides proxy service for remote users. But this approach is little different from the use of other proxy software. It has the disadvantage of being specific to one online vendor, and it still requires the library to maintain, or have access to, a Web server.

As the number of products that libraries wish to offer grows, it becomes less practical to use different validation methods for each vendor. Many online vendors have not yet developed their own solutions to remote validation. They continue to offer either user name and password or IP validation. There is a clear need for standardized methods; for a single doorman watching all the comings and goings. But in the meantime, proxy servers solve both problems. They provide remote access to products that do not offer remote validation, and they provide a common method of validation and remote access for online products from many different vendors. But libraries without IT resources and the means to maintain a proxy server continue to be very limited in the range of remote options they can offer.

Using Extant Patron Lists

The cost, technical skill, and time that are required to manage proxy servers are beyond the means of many libraries. Even those that use proxy servers continue to seek easier and cheaper ways to offer remote access. Improved and standardized methods of direct vendor validation would seem to be the most promising ways to improve remote access for all libraries and lessen the need for proxy servers.

One of the main flaws in current methods of vendor validation is that they do not validate user information from live library patron databases or other existing user ID sources, such as LAN databases or student information systems. Each online vendor has its own method of maintaining separate duplicate lists of user names and passwords, which must then be kept up-to-date. Many libraries also use duplicate lists of library user names and passwords to validate access to their proxy servers. Keeping these multiple password lists is one of the major maintenance tasks associated with proxy servers and remote access.

What we need as an alternative are common methods for vendor servers to validate users on library system and other servers, such as mail or LAN servers, where there are existing user accounts and passwords (Diagram 4). Widely available standardized third-party validation for e-journals and other online products seems a large task, but we are quite close to having this capability and we can encourage vendors to go the final distance to make this possible.

The popular EZproxy software has demonstrated the possibilities for third-party validation by offering methods for validating users on a range of databases such as LDAP (Lightweight Directory Access Protocol), FTP, and POP mail servers. Users connecting to EZproxy are asked for a user name and password. This login information can then be passed to a specified third-party server such as a POP mail server. If a successful login is achieved, indicating a valid account on the third-party server, then EZproxy accepts the user as validated. There is no need for separate username and password lists, or the less secure process of ID pattern matching. For colleges and universities that issue e-mail accounts and other computer accounts to their users, this is a very useful solution that needs to be developed further.

Using What You've Got

The growing use of proxy servers raises a number of important Internet security issues. Server security is already a significant proxy maintenance problem, as last year's Nimda and Code Red[2] and other large-scale attacks on servers have shown us. However, methods for encrypting traffic between proxy servers and different ID servers are already available. Using third-party validation and passing increased amounts of ID and password information over the Internet will present additional challenges, but this may be the impetus we need to for improving Internet security methods.

The "too damned many passwords" problem is a long-standing difficulty. Universities have struggled with ways of reducing the number of separate computer accounts they must maintain and ask their students, staff, and faculty to remember. At my institution, library bar code, student ID, e-mail, and other accounts are still separate. Many institutions have made more progress in using common methods to maintain user information and to validate users on different systems. When we address the need to validate remote library users for many different online products, we are stirring an already-muddy pool.

LDAP (Lightweight Directory Access Protocol) is a well-known standard protocol that presents different databases with a common interchange to information such as common patron information. It is one approach to third-party validation, but it is still not commonly used. Other existing password-protected systems, such as POP e-mail and FTP, present perhaps greater immediate possibilities for third-party validation.

If user validation from existing servers is the way to go, one of the best places to validate from is the library's ILS system itself, a server virtually all libraries have already. Library systems do not provide a means for users to "log in" as mail and network servers do. Integrated library systems do not yet provide a built-in, standardized method for external user bar code validation, but many libraries have customized such validation. At Saint Mary's library we share a GEAC Advance library system with the Novanet provincial library consortium. Some years ago a gateway was customized for this system to validate users for document delivery. User names, bar codes, and telephone numbers are now used to validate users against live patron records. Other librarians are using other systems to validate users in a similar way. So it seems entirely possible to make an interface for third-party validation a standard feature on library systems. If one of the proxy server's roles is that of a friendly but secure doorman, then third-party validation would cut out the middleman, or at least make the job of ID checking a team effort between library ILS systems, proxy servers, and online vendor databases.

The Digital Doorman Is Probably Here to Stay

Online vendors such as ProQuest, Gale, and EBSCO may well add third-party validation to the range of validation methods they offer, as EZproxy has begun to do. I have been told that Gale has plans along these lines for the near future.

More standardized methods of third-party validation, library system validation, and direct vendor validation are likely to become widely available in the next year or two. This will greatly simplify and reduce the cost of remote access as the range of online library products and services grows. It will lessen the need for proxy servers and allow more libraries to offer remote access.

Nonetheless, because they offer common and flexible remote access to such a wide range of online resources, both large and small, proxy servers are unlikely to disappear anytime soon. They will likely continue to develop, becoming easier to operate and offering a greater range of validation options. Proxy servers will continue to be something of an unsung hero in their role as doorman to the still-developing library without walls.

Peter Webster is head of information systems at the Patrick Power Library of Saint Mary's University in Halifax, Nova Scotia, Canada. He holds an M.L.S. from Dalhousie University in Halifax. He has been responsible for online and remote library services at Saint Mary's for the last 10 years. During that time he has been involved with the development of regional online services through the Council of Atlantic Canadian Libraries (CAUL) and Novanet, the Nova Scotia college and university library consortium. His e-mail address is peter.webster@STMARYS.CA.

References

1. "Library Web Proxy Use Survey Results," *Information Technology & Libraries*, Dec., 2001, v. 20, 4, p. 172. Peter Murray's proxy survey is also available on his Web site at http://www.pandc.org/proxy/survey/report.html.
2. Those interested in learning more about the Nimda and Code Red worms and their impact on server security might start with the following: "One Step Ahead," George Hulme; *InformationWeek*, May 20, 2002, p. 57 or http://www.cnn.com/2002/TECH/internet/05/08/nimda.code.red.idg.

To Contact the Companies

Popular Proxy Server Sites:

Innovative Interfaces, Inc.
Web Access Management
5850 Shellmound Way
Emeryville, CA 94608
510/655-6200
(Fax) 510/450-6350
http://www.iii.com/html/customers/c_reference.shtml

Obvia Corp.
Remote Database Access
401 Columbus Ave.
Second Floor
Valhalla, NY 10595
914/773-7859
(Fax) 914/773-7835
http://www.obvia.com

Squid-cache.org
Squid
http://www.squid-cache.org

Useful Utilities
EZproxy
P.O. Box 6271
Peoria, AZ 85385-6271
602/296-0140
(Fax) 888/282-9754
http://www.usefulutilities.com/ezproxy

Selected Database Vendors' Remote Access Information:

EBSCO Publishing
EBSCO*host*
10 Estes St.
P.O. Box 682
Ipswich, MA 01938
978/356-6500
(Fax) 978/356-6565
http://www.ebscoweb.com/SelfHelp/auth.html

Gale
InfoTrac Remote Patron
Authentication Service
27500 Drake Rd.
Farmington Hills, MI 48331
248/699-4253
http://www.galegroup.com/pdf/customer_service/bulletins/rpas.pdf

ProQuest Information and Learning
ProQuest Secured Access
300 N. Zeeb Rd.
P.O. Box 1346
Ann Arbor, MI 48106-1346
734/761-4700
http://www.umi.com/hp/Support/PQD/Secure