Journal of Library Administration
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t792306901

Anything, Anytime, Anywhere

Brian L. Mikesell[a]
[a] St. John's University, USA

To cite this Article Mikesell, Brian L.(2005) 'Anything, Anytime, Anywhere', Journal of Library Administration, 41: 1, 315 — 326
To link to this Article: DOI: 10.1300/J111v41n01_22
URL: http://dx.doi.org/10.1300/J111v41n01_22

Anything, Anytime, Anywhere: Proxy Servers, Shibboleth, and the Dream of the Digital Library

Brian L. Mikesell

St. John's University

SUMMARY. Students and faculty have come to expect off-campus access to the full portfolio of electronic resources made available by their library. They demand, and should be provided, simple access to electronic information sources 24 hours a day, 7 days a week, regardless of their location. Proxy servers have been the solution of choice for remote authentication for some time now, but library users tend to have difficulty with manually configured proxies, and there are beginning to be robust alternatives that can provide secure off-campus access to library resources. Remote authentication should not be a matter of getting a proxy server running and then forgetting about it. New developments should be investigated to ensure the easiest, most reliable and most secure access possible–in the interests of libraries and their users.

KEYWORDS. Technology, Internet, distance education, library services

The dream of the vast, authoritative, easy-to-use virtual library is not only the dream of librarians–our patrons very much share this dream. For some, the ability to access library resources anywhere, anytime is an issue of conve-

[Haworth co-indexing entry note]: "Anything, Anytime, Anywhere: Proxy Servers, Shibboleth, and the Dream of the Digital Library." Mikesell, Brian L. Co-published simultaneously in *Journal of Library Administration* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 41, No. 1/2, 2004, pp. 315-326; and: *The Eleventh Off-Campus Library Services Conference Proceedings* (ed: Patrick B. Mahoney) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2004, pp. 315-326.

http://www.haworthpress.com/web/JLA
Digital Object Identifier: 10.1300/J111v41n01_22

nience and portability. In fact, though, distance learners and others have every right to insist upon its availability not as a convenience, but as a necessity. Increasingly, this dream–and need–is being fulfilled. As more and more publishers and content owners make their materials available via the Internet, library patrons benefit. Libraries have worked to develop the relationships with publishers and vendors that facilitate the provision of electronic information and must continue to ensure that the necessary trust inherent in those relationships is maintained. Part of this process is ensuring that access to these electronic resources is secure and available only to those for whom the licenses are purchased. In libraries and on campuses, this is a relatively simple matter–after all campus networks and workstations can be secured.

When library patrons are not on campus or in the library, though, some method must be employed to both grant them access to the information sources they need as well as to maintain the necessary security. Proxy servers are a widely used solution to the problem of secure remote access. In fact, 60 out of 74 respondents to a survey about proxy server use in 2000 indicated that they were using a proxy server for remote patron authentication (Rogers, 2001, p. 7). As there has not yet been any technology development to supplant proxy servers, there is no reason to assume that the trend has not continued, with even more libraries relying on proxy servers today. A proxy server can provide a legitimate method of remote access, but they are not a complete or perfect solution. A simple, frank discussion of proxy servers and other remote access methods must take place in order to continue to move libraries forward in pursuing the dream of the anything, anytime, anywhere, digital library.

WHAT HAPPENS WHEN LIBRARY USERS ARE ON CAMPUS?

In most academic libraries, patrons may never know that they are using licensed resources that require some form of authentication. This is because libraries have widely adopted IP (Internet Protocol) authentication. "IP validation . . . 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" (Webster, 2002, p. 20). Simply put, IP authentication (or validation) works this way: (1) each organization or institution is assigned a block or blocks of IP addresses–a series of digits that uniquely identifies a computer to the local network and to the Internet at large; (2) the library communicates to its vendors the range of IP addresses used by their institution; (3) the vendor creates on its server(s) a file that says "these IP addresses" are allowed access to these resources; (4) when a student sits down at a computer on campus and clicks on a licensed library resource,

the vendor's server checks the incoming IP address against its file of subscribers' IP addresses; (5) if a match is made, the user is allowed to use the resource, if not, access is denied (see Figure 1).

There are several advantages to IP authentication: users are not required to log in to electronic resources, therefore it is transparent to the users; it is an efficient way to authenticate large numbers of students, faculty, administrators, and staff who want to use library resources online; and it is easy for libraries to administer–all that is required is that the library submit a range of IP addresses to each vendor.

Of course, for users who are not on campus, IP authentication will not work at all. Their IP address is coming from a different source, usually their ISP (Internet Service Provider, such as AOL, RoadRunner, or EarthLink), and when the vendor's server checks their IP address against its list of authorized users, it will not make a match and deny access to the resource (see Figure 2).

THE NEED FOR REMOTE AUTHENTICATION

"In a perfect world, network security wouldn't exist. It's a barrier and, generally speaking, has an inverse relationship with functionality. Librarians and other users of information systems usually find network security to be a nuisance. But just as we understand we should lock our cars when we leave them in parking lots, so we know that we must secure our networks" (Cain, 2003, p. 246). Libraries have worked long and hard to convince publishers that they can safely make their materials available online, and libraries must work to maintain the level of trust necessary to continue the trend. Unfortunately, that means various levels of network security. In addition, "The problem of integrating disparate resources so that they are readily available to the user is both growing and pressing" (Law, forthcoming, para. 16). More and more users are demanding remote access to library resources.

IP authentication is a transparent method of authenticating users to vendors' databases when users are in our libraries or on our campuses. But how do we validate those users who want or need to do research from some other location? There are a range of reasons to want to do this–a graduate student doing

FIGURE 1. Seamless IP Authentication

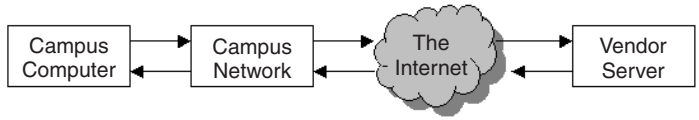
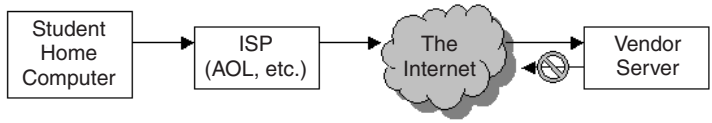


FIGURE 2. IP Address Not Registered–Access Denied



fieldwork; a faculty member at a conference; a student working from home at the end of a day of classes. There are also, of course, library users for whom it is a requirement to have off-campus access to electronic library resources–distance learners; people with disabilities for whom it is difficult to make a trip to campus to do research; students who work full time while taking classes and thus cannot spend much time on campus other than when they are attending class.

No matter what the reason for wanting or needing off-campus access to electronic resources, though, libraries have found various ways of making this possible. Some libraries give out usernames and passwords to particular databases only when requested; other libraries use vendor-developed authentication methods, which require the user to have different accounts for various vendors' databases; perhaps the most widespread method is to use some variety of proxy server.

WHAT IS A PROXY SERVER AND HOW DOES IT WORK?

"The proxy server is not unlike the modern librarian, serving as a helpful and discreet intermediary between users and online information" (Webster, 2002, p. 20). Some users (and librarians) would not agree with this statement. Some types of proxy servers are difficult to use and support, but all can provide an effective means of authenticating remote users into libraries' licensed electronic resources:

> 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 were working from valid IP numbers at the library rather than from their homes or offices. (Webster, 2002, p. 20)

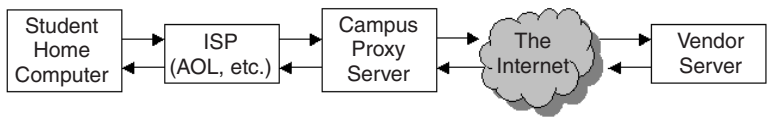
Simply put, a proxy server relays commands from the patron's (off-campus) computer to and from the vendor server. The proxy server is recognized by the vendor's server because the proxy server is on campus and, therefore, has an IP address within the approved range. In "real world" usage: (1) when a library user who is off campus has her browser settings properly configured to use a proxy server (some require manual configuration, others do not) and is connected to the Internet (via her ISP) she can; (2) click on a link to her library's licensed resources; (3) she will be asked to log in to the proxy server; (4) if she has entered a valid username and password, she will be authenticated into the proxy server and can; (5) use any licensed resource that is configured to use IP authentication. In most proxy server installations, the user is logged into the proxy server for the duration of that session–that is, until she closes her browser or specifically logs off the proxy server. This allows the user to browse among the full range of licensed resources made available by the library, regardless of the vendor (see Figure 3).

PROBLEMS WITH PROXIES

An inherent problem with proxies is the issue of security. For example, " . . . any personal computer on a campus network can be set up as a Web server . . . It is also possible to find free proxy server software . . . If the computer is left on, and if a hacker can discover this machine, he has an open door to whatever Web-enabled databases that machine can access" (Cain, 2003, p. 247). It was this kind of "back door" that allowed " . . . an unauthorized user or users exploited unprotected proxy servers from participating JSTOR sites to download illegally more than 51,000 articles from 11 JSTOR journals" (Albanese, 2003, p. 20). This, of course, is not how libraries are using proxies, but relying on a combination of IP authentication and a proxy server makes this kind of abuse possible.

Another aspect of security is authenticating users into the proxy itself. Most vendor licenses authorize only persons currently affiliated with a college or university to access the materials in their databases. This means that it is the li-

FIGURE 3. Proxy Server IP Address Recognized–Access Granted



brary's responsibility to make sure that the list of authorized users against which the proxy server is authenticating remote users is current and valid. Many times, this means periodic extractions of data about current students and faculty from the student information system and subsequent uploads of that data to the proxy server for authentication. Of course, this list must either be updated regularly or risk improperly validating users who should no longer have access (i.e., students who have graduated, etc.) or denying access to users who should be granted access (i.e., newly hired employees, etc.). One way around this particular problem is to authenticate from a "live" source of this data–possibly integrating the proxy server with the student information system, e-mail servers, or some other source that is constantly being updated with the latest data about employees and enrollment. Many libraries, though, lack expertise to make these kinds of links or cannot access these data sources and must rely instead on potentially stale patron data.

One other issue that makes proxies a less than ideal solution is bandwidth consumption. "Remote resources access is the most popular use of proxies in libraries today, but it represents a cumbersome and inefficient way to solve the remote resource access problem. These proxies can be complicated to set up, both for the user and the library, and cause content for the remote resource user to cross an institution's Internet connection twice" (Murray, 2001, p. 176). The fact that off-campus users are logging into a server on campus creates the initial connection to the campus network–they then transmit all of their search commands to the vendor's server via the proxy and the vendor's server then sends data back to the patron via the proxy. This extra layer of data transmission could ultimately create excessive demands on the server and the network. A solution that does not require this extra connection would certainly be preferable.

One final problem with proxy-based solutions is that access is an all-or-nothing proposition. If a user can sign on to the proxy, they can have access to any resource that uses IP authentication. There is no way, really, of allowing access to a particular subset of resources only to a specific group of individuals. This may not be an issue for every library, but getting around it would require a certain amount of creativity and probably extra resources to make the adjustments. One solution might be to run multiple proxies, with separate authentication databases.

As mentioned above, there are a variety of different proxy solutions with different software and relatively widely varying setups. Some libraries choose to use a proxy server that must be configured manually by the user, others use URL-rewriters such as EZproxy, which do not require the user to do any configuration. There are also solutions other than proxies–each with their advan-

tages and disadvantages. Some of these proxy and non-proxy authentication methods are described in the following sections.

REMOTE AUTHENTICATION METHODS

Proxies That Require Users to Configure Browser Settings

A manually configured proxy works basically as above, but users must first alter their browsers' Internet preferences/options so that the browser knows (1) that the user wants to connect to a proxy server and (2) the URL of the proxy server to be used. For example, to configure Internet Explorer 6.6 to use a proxy server, the user would need to: (1) open the browser; (2) click "Tools" in the menu bar; (3) go to "Internet options" in that menu; (4) click on the "Connections" tab; (5) choose the connection they want to configure (usually whatever ISP they use from home); (6) click "Settings"; (7) place a checkmark in the box next to "Use a proxy server for this connection"; and (8) enter the URL and port number of the proxy server they want to use.

To complicate matters, if the user has cable or DSL Internet access, the process is a bit different. Also, the steps are somewhat different for each version of Internet Explorer and Netscape. For an experienced computer/Internet user who has the appropriate instructions for their browser as well as the correct URL and port number, this is a relatively simple matter. For many users, though, this is an enormous hurdle to using library resources from off campus. These users then require additional support from their library. It is possible to write a script that will configure a user's browser for them. This approach has been adopted by some libraries, but requires the programming resources necessary to create and maintain the script properly.

EZproxy

More and more libraries are finding EZproxy to be an easier approach for their patrons. All a user must do is: (1) click on a link to a licensed resource; (2) embedded in such a link is the URL of the EZproxy server, so that; (3) the EZproxy server determines whether the incoming request is from a computer that is on campus or off campus; (4) if the user is on campus, EZproxy steps out of the process and forwards the user to the appropriate database URL; (5) if the user is off campus, EZproxy asks them to log in; (6) if the user inputs a valid username and password, EZproxy will then forward them to the appropriate database URL:

> EZproxy works by dynamically altering the URLs within the web pages provided by your database vendor. The server names within the URLs of these web pages are changed to reflect your EZproxy server instead, causing your users to return to the EZproxy server as they access links on these web pages. The result is a seamless access environment for your users without the need for automatic proxy configuration files. EZproxy only alters references to your database vendors' Web pages, so if your database vendor provides additional links to other free web pages on the Internet, these are left as-is. In this manner, if your users elect to follow one of these links, the EZproxy server is automatically taken out of the communication loop. (EZproxy Overview, n.d., The solution section, para. 2)

There are several advantages to using a URL-rewriter like EZproxy: "(1) users aren't required to make any browser configurations and (2) the proxy server operates transparently, intervening only to authenticate and proxy data for remote patrons" (Bertrand, 2002, p. 135). On the other hand, EZproxy is not without its challenges. For example, the server must have a list of the URLs of resources to which the library subscribes. Creating and keeping this list up-to-date requires steady maintenance.

Onelog

An elegant step forward is a product just beginning to be available in the United States: Onelog, a system produced by ITS Ltd. Onelog might most properly be termed an access management system. At its core, Onelog works in much the same way as EZproxy–that is, it is not invoked until a user clicks on a link to a resource and then it checks to see whether the user is on or off campus, asks for a login as necessary, then forwards the user on to the database. "The Onelog service also features highly advanced IP parsing for offsite access that does not require the user to make any changes to their browser" (Law, forthcoming, para. 9). So, like EZproxy, access to electronic resources is seamless and transparent to the end user.

There are some things that Onelog does in addition, though. For example, "All web-resource scripting is undertaken by ITS as part of the service, thus giving the benison of removing some administrative overheads from the organization" (Law, forthcoming, para. 12). In practice, what this means is that all the library must do is notify ITS (the company that produces Onelog) to which resources they subscribe and ITS takes care of the configuration. Any resource not in their database is added upon request and any access idiosyncrasies worked out by ITS, freeing the library up to do other things. Another big bene-

fit of Onelog is that it can be integrated with one's campus portal and with course management systems. Finally, Onelog can provide the library with extensive statistics about electronic resource usage.

Virtual Private Networks

A virtual private network is " . . . a network that is constructed by using public wires to connect nodes. For example, there are a number of systems that enable you to create networks using the Internet as the medium for transporting data. These systems use encryption and other security mechanisms to ensure that only authorized users can access the network and that the data cannot be intercepted" (VPN, n.d., para. 1). In practice, when a student connects to a virtual private network and, through that, to their library's electronic resources, they are using a more sophisticated version of the proxy server idea. "VPNs extend the institution's IP addresses to machines outside the local area network by tunneling traffic through the general Internet. As such, VPNs work at a network infrastructure layer below that of a Web proxy server, but can accomplish the same result as a Web proxy server for remote resource access" (Murray, 2001, p. 175). The user is still authenticated into the vendor's resource by IP address and still must both have their own ISP and log in to the campus network. The main advantage of a VPN is security. The main disadvantage is the required networking and other technical expertise as well as the hardware and other resources required to set up and maintain the VPN. A VPN may not be the best solution for a library looking to provide remote access to its electronic resources, unless the institution has an interest in providing remote access to their network for other reasons as well.

Athens

Athens, a product of EduServ in the United Kingdom, is a service that manages access to web-based licensed resources by students and other off campus users. "Athens is, fundamentally, a central repository of organisations, usernames and passwords with associated rights. It has extensive account management facilities for organisations to create and manage usernames and passwords, and to allocate rights to individual usernames" (Athens Access Management Services, n.d., Welcome to Athens section, para. 2). Athens is not a proxy-based solution, but rather relies on its centralized system to grant or deny access to a particular resource requested by a user. Athens allows the library to have granular control over who has access to which electronic resources. Also, because

it is centralized, the library is not required to maintain the hardware and software, but only to keep its profile of users and resources up to date.

One item of significant note is that, in addition to the educational institution or library, publishers and vendors must also cooperate with the Athens protocols and be integrated into the system. This, then, requires a level of cooperation beyond merely licensing their materials to libraries. Athens has been in place in the UK since 1996 and, apparently, it is becoming increasingly difficult for a vendor to be successful in the academic library market if they are unwilling to participate in Athens (Athens Access Management Services, n.d., Education section, para. 1).

Shibboleth

A project of Internet 2, Shibboleth is still very much an emerging product. Some limited installations are in place, but it is not ready or available for wider distribution at this time. On its Web site, it is described thus:

> Shibboleth is an initiative to develop an open, standards-based solution to the needs for organizations to exchange information about their users in a secure, and privacy-preserving manner . . . The organizations that may want to exchange information include higher education, their partners, digital content providers, government agencies, etc. The purpose of the exchange is typically to determine if a person using a web browser (e.g., Internet Explorer, Netscape Navigator, Mozilla) has the permissions to access a resource at a target resource based on information such as being a member of an institution or a particular class. (Shibboleth Introduction, n.d., para. 1)

The Shibboleth project was begun after Athens began to show that cooperative methods could be successfully employed for remote access to electronic resources. It does differ from Athens in at least one fundamental respect–it is distributed rather than centralized. Rather than having a centralized depository, an institution must install the software on its own server while also being a member of the Shibboleth community. One similarity to Athens, though, is that publishers and vendors must also participate in the process as members of the Shibboleth community. Shibboleth is based on digital attributes that are exchanged–this is how the "trust relationship" is established between the user's browser and the vendor's server–they must recognize each other. As Shibboleth develops, it may eventually come to replace proxy servers, VPNs, and other remote authentication methods. This is, in part, because it is being

developed with the full awareness of the limitations and difficulties inherent in these methods. Solutions to these problems are being integrated into the end product while it also addresses fundamental issues that cannot even be considered with proxy servers.

DIFFICULTIES WITH NEWER NON-PROXY METHODS

There are, though, certain difficulties with these non-proxy methods of remote authentication. One issue with a solution such as Shibboleth is that many libraries may lack the technical expertise to implement and maintain the technology. Of course, this too may change if Shibboleth becomes a standard, widely implemented solution–especially if provision is made to ensure its simplicity not only for users, but for the institution implementing it. It is likely, though, that there will be libraries–especially smaller ones–that continue to use proxy-based solutions like EZproxy because of the low cost and ease of implementation and use.

The biggest issue, though, is that because of their cooperative nature, there will for some time exist the situation that only a part–whether greater or smaller–of a library's resources will be able to utilize them. For example:

> If the resource is ATHENS protected the user is forced to logon using their ATHENS credentials. This seemed an optimal solution but has proved to be only a partial answer. While it is very satisfactory if all the available resources are ATHENS enabled, it becomes much less convenient if the user intends to move through a variety of resources. (Law, forthcoming, para. 2)

Athens has been available for about seven years, but there are still publishers who do not participate and, thus, libraries using Athens must still have some additional remote authentication procedure for those vendor resources that are not Athens-enabled. The situation is the same, of course, for Shibboleth, especially since it has not yet established itself as a standard solution. As these products continue to mature, it is likely that the vast majority of publishers will come to recognize the benefits of participating, but it is difficult to say when that will ultimately happen.

> Librarians strive to provide transparent systems, with a minimal amount of barriers between the user and the information she or he is seeking. Librarians also want to ensure user privacy and academic freedom. These

> are all laudable values. But the publishing community has legitimate concerns as well . . . To protect our interests, we must protect theirs. (Cain, 2003, p. 247)

Obviously, none of these solutions are perfect, but librarians must continue to investigate new technologies and methods for providing secure, reliable access to the electronic information sources they make available to their patrons. Doing so will help ensure the continuation–and expansion–of the availability of these essential resources while also improving the transparency, integration, and ease-of-use about which library patrons and librarians alike, dream.

REFERENCES

Albanese, A. (January 1, 2003). Open proxy servers victimize JSTOR [Electronic version]. *Library Journal*, *128*(1), 19-20.

Athens access management services. (n.d.). Retrieved December 7, 2003 from Athens Access Management Services Web site: http://www.athensams.net/.

Bertrand, G. (2002). Providing access to remote patrons can be ez [Electronic version]. *Feliciter*, *48*(3), 134-136.

Cain, M. (July 2003). Cybertheft, network security, and the library without walls [Electronic version]. *Journal of Academic Librarianship*, *29*(4), 245-248.

EZproxy overview. (n.d.). Retrieved December 7, 2003 from Useful Utilities Web site: http://www.usefulutilities.com/support/overview.html.

Law, D. (in press). Simplifying access to electronic resources: The changing model of information provision. *The Computer Journal.*

Murray, P. (Dec. 2001). Library web proxy use survey results [Electronic version]. *Information Technology and Libraries*, *10*(4), 172-178.

Rogers, M. (Winter 2001). Proxy servers in wide use [Electronic version]. *Library Journal*, *126*(1), 7. *Shibboleth introduction*. (n.d.). Retrieved December 7, 2003 from Shibboleth Project Web site http://shibboleth.internet2.edu/shib-intro.html.

VPN. (n.d.). Retrieved December 7, 2003 from Webopedia.com Web site: http://www.webopedia.com/TERM/V/VPN.html.

Webster, P. (Sept. 2002). Remote patron validation: Posting a proxy server at the DIGITAL doorway [Electronic version]. *Computers in Libraries*, *22*(8), 18-23.