Improving valid access to site-licensed resources

By Patrick Yott and C. H. Hoebeke

Remote service authentication using a local patron database

If the term *library without walls* no longer circulates among the profession's freshly minted catch phrases, it evokes nevertheless a concept of information service more relevant today than when it was first coined. Among colleges and universities, at least, it is the rare library that does not make its catalog available via local dial-in, or even via the Internet, where it is accessible from virtually anywhere in the world. The phalanx of stand-alone workstations, dedicated to a single CD-ROM or online database and operating on the principle of *one product, one user at a time*, has been largely replaced by networked products that support multiple users simultaneously, and allow access to multiple resources from a single workstation, very often from outside the library confines, in offices and dorm rooms and through dial-up connections.

But as far as we've come, users who have acquired a taste for such convenience have already developed an appetite for more. They are beginning to ask why, if resources can be accessed from the office or dormitory, must there be any geographical limits at all? The faculty member on sabbatical, the graduate student who commutes from outside the local calling district but resides near a public library with Internet access, or any of the valid users who purchase private Internet accounts are beginning to wonder why access from outside the college or university is typically limited to the basic catalog. What about databases, dictionaries and encyclopedias, and the scores of informational Web sites that can be accessed on or by dialing into campus?

The reason, of course, is that library site licenses for these resources are usually protected by restricting access on the basis of the institution's IP domain. Every computer linked to the Internet has a unique IP address. Depending on local policies and configurations, the addresses might be permanently assigned to each machine or temporarily assigned for the particular session in which the computer is connected to the Internet. Either way, all of these unique numerical addresses fall within a range or group of ranges that is assigned to the college or university. When a library purchases a site license, it gives the vendor a range of valid IP addresses in the campus domain. Users who try to access an IP-restricted resource from outside the domain—for example, from CompuServe or America Online—will be denied access, *even if they happen to be faculty or students of the institution that has purchased the site license.* In other words, though the perimeter has expanded, the library walls have by no means been removed.

Approaches to authentication

A number of institutions have been nudging vendors to come up with alternative means of validating users, or have devised their own. One method is to establish a go-between server that requires the user to log in with ID and password, and that, in effect, spoofs the vendor's computer into thinking that the remote user is actually in the valid domain. Another solution has been the implementation of Z39.50 systems allowing users to log in to a

Patrick Yott is coordinator of social sciences data services at the University of Virginia Library; e-mail: pmy2n@poe.acc.virginia.edu. C. H. Hoebeke is assistant director, interlibrary services, at the University of Virginia; e-mail: chh4n@server1.mail.virginia.edu

client machine that, upon successful log-in, will send a hidden ID and password to the vendor's system.

Both of these solutions require considerable computing sophistication and investment, and have their limitations. The go-between server can be slow because every transaction between the user and the remote service must travel through an intermediary computer. Z39.50 requires that the desired resource have a Z39.50 server, and while this solution has the advantage of allowing libraries to provide a common interface for all databases that are mapped to the Z39.50 client, achieving the "common denominator" often entails the sacrifice of power search techniques available in the database's native interface. And because Z39.50 is based on the MARC format, it does not lend itself to searching and displaying full text.

A third option that has been devised in getting around the restraints of IP restrictions is a simple cgi script that validates users against a local database; for example, that maintained by the registrar and personnel offices. At the University of Virginia, numerous Web forms take advantage of such a database, authenticating log-ins and passing on needed personal information so that users can make Web-based interlibrary loan requests, suggest purchases, and, in a few instances, access site-licensed databases. This approach has been used for nearly a year now by our health sciences library in providing access to OVID, and has been recently adapted for use with the university's UnCover gateway.

This method gets more complicated in cases where only a certain portion of users—faculty and graduate students in specific departments, for example—are allowed access to a remote resource. In this scenario, the standard IP protection is seriously flawed. Not only would it prevent the professor on sabbatical from having access to a service to which he or she is entitled, it would also permit access to any person off the street, *solely because that person happened to be sitting at a terminal with a valid IP address.*

The challenge in this scenario is not only to verify institutional affiliation, but to restrict usage to an authorized subset of the library's normal user population. One solution in recent years has been to provide the vendor with a file of all legitimate users, and let the vendor validate the patron on the host end. Since this model necessitates frequent updating, new students and employees will more than likely incur delays, depending on how often the vendor is willing to make updates and on how often the library or systems staff can take the time to extract and submit the updated data.

Furthermore, submitting patron files to vendors raises serious privacy issues. Any kind of automated validation requires a "unique ID," and one that is reasonably private. For convenience many institutions rely on the Social Security number, or some modified version of it, for all manner of automation tasks, since it is guaranteed to be unique and, in theory, at least, is known only to the user and to those staff who have a need to know.

Unfortunately, providing someone's Social Security number to a third party without consent is of dubious legality. But even if the patron file does not contain Social Security numbers, there is still a privacy issue. The fact that a student or faculty member is eligible for a particular service does not mean he or she will want to partake of it. A library is not therefore warranted in releasing, en masse, the personal data of its users.

On both counts, then, timeliness and privacy, local validation has the advantage over submitting patron files to the vendor, because authentication is performed against a database that is updated on the spot as the need arises, and because whatever personal information is used is explicitly or implicitly authorized by the patron when he or she initiates the log-in to use the service.

Down, dirty, and effective—one library's answer

In the following example, members of a valid subset of the university population are authenticated against a local database, a whois server maintained by the university's central computing division, then logged in to a site-licensed service, without in any way limiting access on the basis of IP. A Web form prompts users to provide both their last name and university ID. Forms submitted without both pieces of information automatically cause the authentication to fail. Both pieces of data are required in order to provide a modicum of security and to eliminate the accidental match on a common name, such as Jones or Smith. Once the program has received both pieces of data, it runs that data through our local whois server, and based upon the result of that lookup, processes the user accordingly.

Executing a whois lookup from the prompt yields the following information (this is the result of configuration at the University of Virginia—your results may vary):

```
Name:             Patrick M. Yott
Mailid/Handle:    pmy2n
Unix Uid:         41335
Classification:   Faculty
Department:       Ald Lib-Social Sci. Svcs.
Office Phone:     (804) 982-2630
Registered Addr.: pmy2n@Virginia.EDU
                  pmy2n@poe.acc.virginia.edu
```

The script used for this validation is written in PERL, and it is quite likely that there are individuals at your campus (if not your library) that are currently writing PERL code.

Once the validation script has determined that both pieces of information have been provided, it executes a whois lookup and reads the resulting information into an array (list). Each line returned by the whois lookup is an item (element) of that array. We then evaluate various items to confirm the user's status and university affiliation (he or she must be a UVA faculty member or graduate student not associated with a professional school). An element of an array is identified by its place in the list (subscript value). Therefore, to evaluate the line containing classification we examine the fourth item in the list (subscript 3), and to evaluate departmental affiliation we examine the fifth item in the list (subscript 4).

To pass this validation test, the fourth item in our list ($array[3]) must contain (=~) either the term *Faculty* or *Grad*, and the fifth item must not contain (!~) any of the professional schools.

Let's assume that our user has passed the whois validation. At this point our script returns HTML code to the client's browser, resulting in a blank page with a submit button. By clicking on that button, the now authenticated user is taken to the remote service, and is correctly logged in to the restricted system.

If our user supplied both pieces of information, but failed the whois lookup, we need to return some text explaining the problem. Again, our script returns some rudimentary HTML to the client's browser.

Finally, we need to return a polite admonishment for those who failed to provide both pieces of the required information.

This particular script, with slight modifications, has any number of applications, and indeed, Web forms with input boxes for "last name" and "University ID" are appearing on more and more Web pages throughout our campus. Naturally, the technical resources at hand will largely determine to what extent the validation scheme used by the University of Virginia can be employed by other institutions.

Conclusion

In terms of the larger issue of local validation, neither the programming language nor the type of user database is of critical importance. What is important is that the database is accessible via the Internet, that it contains data that can be used to include or exclude users on whatever criteria are called for, and that it contains a unique identification value for each individual, a value that is known to the user but is not available to the general public.

Of at least equal importance is the cooperation of vendors, too many of whom protect their site licenses on the basis of IP. Until more libraries insist on the liberation offered by local validation, the walls remain. ■

(Benchmarking cont. from page 694)

spent at the desk by "extra" librarians (i.e., not just the two who were scheduled) and the difference between the number of questions recorded on the statistics sheets and the number of users who come to the desk. The extra librarians contributed 10 percent of the time that was recorded; 15 percent more people were helped at the desk than were recorded on the statistics sheets.

The study's greatest value has been to give us figures that we can use to bolster our arguments for maintaining at least current levels of staff.

Notes

[1] Gerald L. Balm, *Benchmarking: A Practitioner's Guide for Becoming and Staying Best of the Best* (Schaumburg, Ill.: QPMA Press, 1992), 16.

[2] Joanne G. Marshall and Holly Shipp Buchanan, "Benchmarking Reference Services: An Introduction," *Medical Reference Services Quarterly* 14, no. 3 (1995):59–73.

[3] Susan Greco and Chris Caggiano, "Software Support: Please Hold for Customer Service," *Inc.* 13, no. 8 (1991):96.

[4] Karen L. Katz, Blair M. Larson, and Richard C. Larson, "Prescription for the Waiting-in-Line Blues: Entertain, Enlighten, and Engage," *Sloan Management Review* 32, no. 2 (1991):44–53. ■