



User Authentication

A SPEC Kit compiled by

Terry Plum
Assistant Professor
Simmons Graduate School of Library and Information Science

Richard Bleiler
Humanities Reference Librarian
University of Connecticut

December 2001

Series Editor: Lee Anne George

SPEC Kits are published by the

Association of Research Libraries
OFFICE OF LEADERSHIP AND MANAGEMENT SERVICES
21 Dupont Circle, NW, Suite 800
Washington, D.C. 20036-1118
(202) 296-2296 Fax (202) 872-0884
<<http://www.arl.org/olms/infosvcs.html>>
<pubs@arl.org>

ISSN 0160 3582

Copyright © 2001

The papers in this compilation are copyrighted by the Association of Research Libraries. ARL grants blanket permission to reproduce and distribute copies of these works for nonprofit, educational, or library purposes, provided that copies are distributed at or below cost, and that ARL, the source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the U.S. Copyright Act.



The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

SPEC

SUPPORTING EFFECTIVE LIBRARY MANAGEMENT FOR OVER TWENTY YEARS

Committed to assisting research and academic libraries in the continuous improvement of management systems, OLMS has worked since 1970 to gather and disseminate the best practices for library needs. As part of its commitment, OLMS maintains an active publications program best known for its SPEC Kits. Through the OLMS Collaborative Research/Writing Program, librarians work with ARL staff to design SPEC surveys and write publications. Originally established as an information source for ARL member libraries, the SPEC series has grown to serve the needs of the library community worldwide.

WHAT ARE SPEC KITS?

Published six times per year, SPEC Kits contain the most valuable, up-to-date information on the latest issues of concern to libraries and librarians today. They are the result of a systematic survey of ARL member libraries on a particular topic related to current practice in the field. Each SPEC Kit contains an executive summary of the survey results (previously printed as the SPEC Flyer); survey questions with tallies and selected comments; the best representative documents from survey participants, such as policies, procedures, handbooks, guidelines, websites, records, brochures, and statements; and a selected reading list—both in print and online sources—containing the most current literature available on the topic for further study.

SUBSCRIBE TO SPEC

Subscribers tell us that the information contained in SPEC Kits is valuable to a variety of users, both inside and outside the library. SPEC purchasers use the documentation found in SPEC Kits as a point of departure for research and problem solving because they lend immediate authority to proposals and set standards for designing programs or writing procedure statements. SPEC Kits also function as an important reference tool for library administrators, staff, students, and professionals in allied disciplines who may not have access to this kind of information.

SPEC Kits can be ordered directly from the ARL Publications Distribution Center. To order, call (301) 362-8196, fax (301) 206-9789, email <pubs@arl.org>, or go to <<http://www.arl.org/pubscat/index.html>>.

Information on SPEC and other OLMS products and services can be found on the ARL website at <<http://www.arl.org/olms/infosvcs.html>>. The website for SPEC is <<http://www.arl.org/spec/index.html>>. The executive summary or flyer for each kit after December 1993 can be accessed free of charge at the SPEC website.



Kit 267

User Authentication
December 2001

Survey

EXECUTIVE SUMMARY	9
SURVEY RESULTS	14
RESPONDING INSTITUTIONS.....	22

Representative Documents

EXPLANATIONS FOR USERS

University of Alberta

<i>Off Campus Access to Electronic Library Resources.....</i>	26
---	----

University of California–Davis

<i>Access to UC Davis Licensed Resources</i>	27
--	----

University of California–Irvine

<i>Connecting from Home: Remote Access Guide</i>	29
--	----

University of Colorado

<i>Remote Access to Chinook: CU–Boulder Students, Staff, and Faculty</i>	31
--	----

Colorado State University

<i>Remote Access to Library Databases: Proxy Server Instructions</i>	33
--	----

<i>How the Proxy Server Works</i>	34
---	----

<i>Without the Proxy Server</i>	35
---------------------------------------	----

<i>With the Proxy Server</i>	36
------------------------------------	----

University of Connecticut

<i>Obtaining an Account on the University of Connecticut Proxy Server.....</i>	37
--	----

Cornell University

<i>Proxy Server for Access to Library Resources from Outside Cornell.....</i>	39
---	----

Dartmouth College	
<i>Kerberos Authentication at Dartmouth</i>	41
<i>Kerberos Authentication. How it Works</i>	42
Duke University	
<i>Remote Access Options. Connecting from Off-campus</i>	43
University of Florida	
<i>Guide to Remote Access to Library Databases</i>	44
<i>Guide to Remote Access Using a Proxy Connection</i>	46
McMaster University	
<i>Off Campus Access: A Guide to the McMaster Proxy Service</i>	48
University of Manitoba	
<i>Using the Proxy Server to Access Restricted Databases and Web Resources</i>	49
Massachusetts Institute of Technology	
<i>Obtaining MIT Certificates: Quick Guide</i>	51
University of North Carolina	
<i>Off-campus Access via Proxy Server</i>	52
University of Oregon	
<i>Off-campus Access to Library Databases</i>	55
University of Washington	
<i>About UW NetIDs</i>	56
<i>Connecting to the Libraries</i>	57
Washington University	
<i>Proxy Setup Instructions</i>	59
University of Waterloo	
<i>Connect from Home</i>	61
TECHNICAL SPECIFICATIONS	
Cornell University	
<i>How the Proxy Web Server Works</i>	64
Université Laval	
<i>Serveur Mandataire. Version 1.0. Détails Techniques</i>	66

PROJECT PLANS

University of Connecticut

University ITS. Authentication Project 84

Authentication Project Profile 85

University of Waterloo

Identification, Authentication, and Electronic Commerce 94

Final Report for the ID-AUTH-ECOMM Prototype. Preface 95

Final Report. Introduction..... 96

Final Report. Recommendations..... 98

Selected Resources

BOOKS AND JOURNAL ARTICLES 101



SURVEY



Executive Summary

Introduction

Until the advent of the World Wide Web and the concomitant development of global computer networks, most research libraries could provide access to their resources with few concerns about the status of those who sought the information, or concerns that the information was restricted to certain classes of users. Developments in computer technologies have irrevocably altered library operations, and it is now the exceptional library that has not in some way responded to the challenges of authenticating and authorizing its users, particularly those users needing to access the library's systems and networked information resources from remote locations. Furthermore, licenses for networked information resources increasingly require authentication controls and need to specify different levels of authorization.

Authentication and authorization are complex access management processes that involve the verification of the identity and status of the user, the ways in which IP ranges are limited, the processes by which information technology support is handled, and the systems by which authentication information and authorization are maintained.

This SPEC survey is designed to examine the systems that research libraries use to authenticate and authorize the users of their online networked information resources. For the purposes of this survey, authentication is defined as the process of determining whether someone or something is, in fact, who or what he declares himself to be. Authorization is the process of giving someone permission to do or have something, including privileges of use, such as access to file directories, amount of allocated storage space, access to licensed electronic resources, and so forth. Networked information resources are defined as electronically

accessible information resources (e.g., library or academically developed databases, university databases, commercial databases, full-text services, e-journals, etc.) funded or enabled by the library, which are made available to authorized users through an intentional and systematic network (LAN, WAN, dial-in, etc.).

This survey was distributed to the 121 ARL member libraries in spring 2001. Fifty-two libraries (43%) responded to the survey.

Authentication

Fifty-one of the responding institutions (98%) stated that they authenticated their users in some way. The one library reporting no authentication apparently limits by IP, has a proxy server, and offers remote access services to library resources through a modem pool. Therefore, all of the responding libraries authenticate users of networked resources.

In selecting user categories for authentication, 48 respondents (96%) authenticate staff, and 46 (92%) authenticate their undergraduate students, graduate students, and faculty. Seven libraries (14%) authenticate alumni and local community members, although their access to the library's networked electronic resources is apparently much reduced. Fourteen respondents (28%) report that they authenticate other categories, providing access to groups such as Friends of the Library, extension faculty and students, selected department-sponsored guest accounts, university affiliates, affiliated institutions, and external clients, including international clients.

There were 49 responses (96%) to the question about the number of networked information resources made available to authenticated users, but the numbers varied enormously, from 1 to 23,806. That these numbers are so disparate almost certainly

demonstrates that there is no agreement on the unit of analysis for measuring “networked information resources.” A simple example will suffice in showing the nature of the problem: does JSTOR count as a single networked information resource, or is it the sum of its subscription modules, or is it the sum of the journal contents? And, if the latter, how are title changes treated? This problem is only exacerbated when one considers the number and variety of networked electronic resources potentially available to ARL member libraries.

Access Management Systems

There was equal divagation in the responses on the type of access management system used to authenticate users of networked information resources, although in this case the disparity appears to be the result of respondents using multiple systems. For example, 42 of the 51 responding libraries (82%) report using IP addresses to authenticate users of at least some portion of their networked electronic resources, while an overlapping 40 (78%) report the use of password and user ID. There is similar overlap among and within the other types of systems. (See question 4 in the Survey Results section for a chart of the types of systems used.)

Surprisingly few libraries (10, or 20%) use the ILS to authenticate patrons, and only four (8%) use a non-ILS gateway, such as OCLC’s WebZ. Based on a review of the literature and websites, the number of libraries that use Public Key Infrastructure (PKI) is also unexpectedly small. Since only three libraries (6%) claim that they have enabled this scheme to authenticate patrons, perhaps PKI is as yet more discussed than implemented.

All together, 46 respondents report using some kind of proxy server. These libraries apparently are migrating from mechanical proxy systems (pac files) to application-level proxy servers or rewriters (such as EZproxy). Twenty-six libraries report using a mechanical proxy system, whereas 12 use an application-level rewriter. Two libraries use both. Fifteen respondents (33%) use EZproxy as the proxy server software, almost twice as many as use Apache (9 respondents or 20%), Squid (8, or 17%), or Web

Access Management (III) (6, or 13%). Only three respondents use Netscape Iplanet Proxy, and one uses Microsoft Proxy. A small number of institutions are using homegrown or custom solutions. Of the 42 libraries that rely upon IP authentication, only four have neither a proxy server nor a modem pool, while only two still rely upon a dial-in modem pool exclusively for off-campus access.

Survey respondents were asked which database of patron information is used to verify eligible users. Again, it is evident that multiple systems are in use at many institutions. In the majority of instances (28 responses or 55%) some portion of the authentication system checks against a dynamic patron load or circulation patron record database with the ILS. Fewer institutions (20, or 39%) have a system that checks against a dynamic institutional personnel database. An equal number (14, or 28%) use a system check against a flat file extracted from the ILS or a system check against a separately created database of eligible users. Only seven (14%) use a system check against a flat file extracted from the institutional personnel database. Five respondents (10%) use other solutions, including a Virtual Personal Network and personal certificates.

Authorization

What services are being authorized after authentication? The access provided by credential-based (passwords, certificates, etc.) systems is widely distributed. At 35 institutions (69%), authenticated users can access their personal circulation record, while at 33 (65%) users can request interlibrary loans and document delivery and use e-mail. Other accessible services include holds and recalls on books (31 institutions or 61%), course registration information (29, or 57%), databases (28, or 55%), and both course reserve materials and file space on the network server (25, or 49%). Numbers were lower for accessing financial records and computer labs (21, or 41%) and lower still for accessing e-books and e-journals (20, or 39%). The figures continued to drop for accessing the library OPAC (19, or 37%), distance education courses (18, or 35%), and transcripts (17, or 33%). Only seven institutions (14%) use credential-based systems to provide access to photocopy

machines, dining facilities, and groupware.

The distribution of services provided by proxy systems and IP filtering is almost identical. The majority of institutions that use a proxy server provide access to databases (41, or 80%), e-journals (40, or 78%), and e-books (36, or 71%). Fourteen (27%) provide access to course reserve materials, and 11 (22%) to the library OPAC. There are few uses of the proxy server for other purposes, although six respondents (12%) report using it for distance education courses.

User Privacy

Questions of confidentiality and privacy policy are important to any web-based authentication system. Although this survey did not specifically inquire about privacy policies, it surveyed respondents about how user information is tied to the respective search session and whether such information is archived. Thirty-two of 49 respondents (65%) provide *anonymous access*, in which each session is anonymous and repeat users cannot be identified. Twenty-four (49%) provide *identified access*, in which actual identities are associated only with sessions, and 12 (25%) provide *pseudonymous access*, in which repeat users can be identified, but the identity of a specific user cannot be determined. These responses appear to be closely allied with the American Library Association (ALA) *Policy Concerning Confidentiality of Personally Identifiable Information about Library Users*. This policy states in part that, "Confidentiality extends to 'information sought or received, and materials consulted, borrowed or acquired,' and includes database search records, reference interviews, circulation records, interlibrary loan records, and other personally identifiable uses of library materials, facilities, or services." Nevertheless, ten respondents (20%) archive identified access data and eight (16%) provide pseudonymous access with demographic information that does not provide actual identities. The ALA *Policy on Confidentiality of Library Records* strongly recommends that, "Responsible officers of each library, cooperative system, and consortium in the United States formally adopt a policy which specifically recognizes its

circulation records and other records identifying the name of library users to be confidential in nature." It appears that most libraries address this issue by refusing to store ID information with session data.

System Management

Who does the work of building and maintaining the authentication system? In general, much of the work is done within the library. Central library information technology (IT) staff manage the authentication and authorization systems in 46 of the responding institutions (90%). Institution or campus-wide IT staff manage them for 37 (73%) of the respondents and the vendor of the networked information resource manages authentication for 25 (49%). One respondent wrote that campus IT handles e-mail, registration, and distance education; the vendor handles IP authentication for e-journals, e-books, and online databases; and the library handles the other resources. This is a very traditional arrangement.

Question 9 asked how many IP-filtered resources are managed at the vendor, consortium, institution, or other level. The data again illustrate the difficulty of counting networked information resources with responses ranging from zero to more than 10,000. A general impression is that the vendor most often handles IP-based access restrictions to library resources.

At the 26 institutions that manage access by an ILS capable of accessing databases through Z39.50 (z-client), survey results indicate a fairly equal distribution across ILS vendors. Endeavor and III are each used by five of the 26 respondents (19%), SIRSI is used by four (15%), and DRA is used by three (12%). Six respondents (23%) use other vendors, which included Geac Advance, CDL and Melvyl, and EpixTech Horizon.

When access is managed by passwords and user IDs, passwords are typically randomly generated and are not expired on a regular schedule. The 40 responses to the question on how often passwords are expired showed evident confusion about the process. In several instances, there was no mechanism in place to expire passwords, and they thus did not expire or expired "rarely" or with "no

set period," although one respondent indicated that this was under review. Several respondents indicated that passwords expired at a fixed time—at the end of each quarter, semester, term, or annual session—but several also indicated that expirations were dependent on the status of the individual, and that expirations occurred when the person graduated or left the University.

There are a variety of schemes in place to prevent the hijacking of user IDs and passwords by unauthorized persons, including SSL and IP restriction to webpages that list passwords. There are also various methods for authentication when the initial user ID and password are assigned: for example, in-person registration with photo ID. Nonetheless, there appear to be no technical schemes for limiting the distribution of user IDs and passwords by authorized users to unauthorized users. Most institutions that recognized the existence of this problem have a use policy or honor system.

There are few respondents who manage access through a library gateway such as WebZ. Four of the 13 respondents (31%) use OCLC products such as WebZ and SiteSearch. The others use a webscript from OCLC, DRA Web2, VTLS, WebVoyage, or Voyager. Slightly fewer than half of this group has a gateway that provides access to both Z39.50 and non-Z39.50 resources. Only five respondents report that they manage access by digital certificates. Two of these use Kerberos and two use VeriSign.

The future of access management systems appears to indicate some turmoil. Of 47 respondents, almost half (22, or 47%) indicated that they plan to switch to a different access management system within the next two years. The responses are heterogeneous and most respondents seem still to be in the planning stage. The new systems most often mentioned are EZproxy, digital certificates, and LDAP. Several respondents indicated that they had identified nothing, but that evaluations were being done. These changes would bear reexamination.

Conclusions

One of the inescapable conclusions to emerge from this survey is that research libraries do not appear to possess a common standard or

a common vocabulary that can be used for measuring, describing, and communicating their holdings of networked information resources. This is evident from the disparate responses to the deceptively simple question involving the number of networked information resources made available to authenticated users. That the responses are so disparate certainly demonstrates that there is no agreement on the standard unit of analysis for measuring the concept of "networked information resource." Libraries that can provide statistics on books and serials have more difficulty counting electronic resources.¹

Another conclusion emerges directly from the library responses to the application-level proxy rewriter. Those libraries that have implemented an application-level proxy rewriter are able to serve additional resources through it, offering more resources than those libraries that do not use the application-level proxy rewriter. The application-level proxy server occupies a more central position in the authentication system than mechanical proxy servers. Although relatively few users rely upon a credential-based system to provide network access and authentication, an optimum system might integrate a credential-based system and the application-level proxy rewriter, eliminating the need for a gateway access management system.

Apparently, the Z39.50 standard no longer plays a significant role in the way research libraries provide access to their networked information resources. When only 20% of the responding libraries use a Z39.50 capable ILS for authentication and when that authentication represents only 6% of authenticated usage in those libraries, it is safe to say that the standard is less relevant than it once was and is no longer the future for libraries in meeting the information needs of their users.

Confidentiality of library records appears to be addressed by the deletion of ID data from session information. Most libraries are addressing the *ALA Policy on Confidentiality of Library Records* by establishing a system where the ID data cannot be retrieved because they are not archived. We think this phenomenon deserves further attention, and speculate that the ID data are in fact valuable and

should be retained. It is perhaps a lack of trust in policies or a lack of confidence in technical security that has encouraged so many institutions to remove this data.

There was little homogeneity in the 37 responses to Question 14, which involved security measures to insure that passwords and IDs are not distributed inappropriately. Some respondents use manual (personal) ID checks and verifications; others attempt to create secure pages with special logins and PINs; still others attempt to link to databases (payroll, registration, human resources) and to verify validity with other systems before establishing an account. Passwords likewise are distributed in all possible ways, with some respondents offering immediate validation while others send personal mail containing the passwords. Once the passwords are distributed, most institutions rely upon honor codes and signed use agreements to limit further distribution of the passwords by authorized users to unauthorized users. That there is such little agreement in the ways in which authorized and unauthorized users can be determined would seem to necessitate additional research and further study of this subject.

Finally, there were a number of comments concerning the heterogeneity of the authentication systems. "Extremely heterogeneous network," "There are a lot of different authentication and authorization systems all around campus," "Question 4 was difficult to answer since we often use a combination of methods for our authentication routines." And most tellingly, "We have three layers of authentication." There are both layered, integrated systems and those with different authentication systems for different services. The complicated authentication environment is a difficult picture to survey. It does seem clear that, at many libraries, system analysis is being applied to authentication schema, and efforts are being made to integrate disparate authentication systems into a layered or sequential approach.

¹ Editor's note: The E-Metrics project, one of the ARL New Measures Initiatives, is an effort to explore the feasibility of defining and collecting data on the use and value of electronic resources. Information about the project is available on the ARL website at <http://www.arl.org/stats/newmeas/emetrics/index.html>.