DLPS Security System Specifications

Abstract

For several years, DLPS and DLPS-precursor organizations have recognized the need for a method of authentication and authorization with more precision than IP addresses.[1] Discussions with ITD and past efforts have focused incorrectly on the Wolverine Access system as a model for authentication and authorization of library resources. Wolverine Access relies on the user having a current Kerberos identity in the umich.edu realm (essentially, presence/absence in X.500),[2] and SSL to encrypt the data stream. DLPS needs a system for authorizing access to library resources where the user population differs significantly from the user population represented in the umich.edu realm. Moreover, DLPS requires that the system only encrypt the authentication process and that the flow of data *not* be encrypted. This paper puts forward a recommendation for establishing and maintaining a database of users of online library materials managed by DLPS, with methods for authentication and communication consistent with Library needs.

Assumptions

1. DLPS needs to address authentication and authorization for campus-based resources managed by the DLPS. While there are needs for remotely managed resources (e.g., *Encyclopedia Britannica* and Lexis-Nexis), and the solution recommended here may aid in meeting those needs, this set of recommendations is intended to address only resources managed locally.

---

[1] While IP addresses are insufficient for many purposes, they remain centrally important for some types of authentication. Each University Library contract for an information resource seeks to include language supporting all in-library use, without individual user authentication (but instead permitting access by Library IP addresses). The historical role of the Library as a resource for the community and for unaffiliated scholars (especially, for example, from smaller colleges and universities in the region) is an important one that the Library seeks to sustain in the electronic environment. Consequently, we continue to see IP addresses as one key mechanism for authentication. For more information, please see Appendix 1: Context

[2] X.500 is a useful mental representation for the umich.edu realm, but they are technically and administratively separate, using entirely different databases. I have assumed that there are no valid uniquenames that are also **not** in X.500, but this is not entirely correct. In querying Sarr Blumson about this, Sarr said: “This is an ‘administrative goal’ but it isn't achieved; I believe that among the cloud of people from Engineering who have PEAK hassles because they don't get ‘@umich.edu’ mail there have been a few who weren't in X.500 at all. There have also been a few who were in X.500 twice. Some of the processes that keep then consistent are manual. There is certainly no technical reason why you can't be in one but not the other.”

2. There is a poor correlation between current users in the umich.edu realm and the user population for online library resources.

    a. As many as 40% of the users in the umich.edu realm are in the database because they have not been purged since the individual's departure from the University.

    b. The umich.edu realm does not contain a body of users who are significant and legitimate library resource users (e.g., doctoral candidates who are not currently enrolled in classes).

    c. The KA (i.e., Kerberos Authentication) Server for the umich.edu realm lacks a means to register external users for services such as those hosted by DLPS for the University of Michigan Press and the Association of Asian Studies; DLPS resources require us to maintain a database of those users and to permit them to have access to selected resources.

3. There is a low likelihood of adjusting/adapting the umich.edu realm's administrative practices to accommodate Library authorization/authentication needs.

    a. The purposes of the umich.edu realm (i.e., primarily of identifying current users of ITD resources) and the proposed database of library service users are different.

    b. The administrative practices of the KA Server for the umich.edu realm are not consistent with the needs of a library authorization/authentication mechanism (e.g., it is purged too infrequently, and updates are not added frequently enough). By contrast, the Library Systems Office (LSO) updates parts of its user database weekly[3]

---

[3] Sarr Blumson notes: “I'd worry about this claim. The reason that X.500 and the Kerberos database don't get purged is that it's difficult to tell from the data that DSC sends that someone has left, and dangerous to assume that because they're not on the new tape that they're really gone. I'd want data to back up my claim that the LSO actually does do better.”

Phyllis Valentine, Interim Director of the Library Systems Office responds: “Faculty/staff data is received and loaded once a month; the records received correspond with our definition of eligible patrons, looking at appointment level and/or percentage of effort. Student data is received and loaded generally once a month except for Sept. and Jan., when there are usually two loads. We receive records only for currently enrolled students. Individual records in both groups can and are updated on a case by case basis daily. Special patron records are added to a different patron region and may follow different rules, but all are added and edited manually. The load process sets an expiry date for each patron record, as well as loading information about campus and level (fac, staff, undg, grad, geo, etc.). [The Circ committee chairs ... set the load schedule and the base/expiry dates that are set for each load. In general the expiry date is set for about a week after the next load date, which allows for some slippage in loads...and prevents folks from being blocked too soon. In summer, to prevent patrons from being expired, the date is set longer.] So lack of a record for an existing patron effectively turns off their eligibility. We use this information

4. Any approach must take into account the Library's need to encrypt the authentication process and yet not encrypt the flow of data.

    a. Kerberos identities must be kept secure; consequently, this data stream must be encrypted. Clearly, a true Kerberos authentication is preferred over sending the user's password over SSL, but this has been explored by ITD and dismissed as impractical.

    b. All user password information should be kept secure, and thus should be encrypted.

    c. While some services, e.g., Wolverine Access, transmit private and sensitive data such as credit card numbers, and so should be kept secure, Library services transmit licensed public data, frequently in large volume (e.g., multi-page documents or high-resolution images), that need not be kept secure. (That is, we need only make "reasonable efforts" to ensure that only the authorized user has access; reasonable efforts have only included authorization, and not secure lines.) Because of the overhead of SSL, performance and reliability issues associated with it, and the public nature of this data, we require that the information transactions themselves *must not* be encrypted.

5. Current solutions may differ from future solutions.

    a. We believe that a Public Key Infrastructure (PKI) is the answer to universal authentication and authorization needs in the area of licensed information goods.

    b. Cookies and annotated URLs are the current answer to authorization.[4]

    c. Kerberos is an imperfect answer to authentication because there is a poor correlation between current users in the umich.edu realm and the user population for online library resources (see #2, above).

    d. Users will frequently have access to *some* Library resources, but not *all* Library resources. For example, a non-affiliated user, through a subscription agreement with the University of Michigan Press, may have access to the Old English Corpus and the Middle English Compendium, but no other DLPS collections. Kerberos has a limited capacity for expressing these roles (i.e., only PTS groups, and not attributes of the user identity in the authentication process);

---

to provide or refuse access to library privileges and MIRLYN files. We have purged patron files, but leave in records for patrons who still have library materials and/or fines; these patrons do not have privileges however."

[4] Blumson notes: "'Wolverine Access' (the mechanisms are widely used at this point) actually uses a compromise system. The cookie (they use cookies, but it could be annotated URLs) is only a key to a semi-persistent (it expires) object on the server, which holds the actual Kerberos tickets."

consequently, the managers of those Library resources must maintain a database of privileges (i.e., rights of access for specific individuals).

e. As implied in 5d, above, the Library must support limited and selective access for users and groups outside of the University of Michigan.

Specifications

The proposed system will be built using Oracle tables and will be managed by DLPS, with data supplied by the Library Systems Office (LSO). DLPS and LSO staff, along with designated representatives from selected collection providers (e.g., a staff member from HRAF) will have maintenance privileges, and the database will otherwise be publicly inaccessible. The system will use either cookies or annotated URLs to provide authorization data to the browser, with no preference being given at this time to either method. Users will be able to use a password and user identity specific to and generated by this system; additionally, UM users may choose to use Kerberos identities and passwords to authenticate.

Database of users and collections

Databases of users and collections will be maintained by DLPS. The user database will determine if the user has a valid identity in the system, and, by virtue of a password mechanism, whether the user is who s/he claims to be (i.e., authentication). Users will have access to some or all of the collections maintained by DLPS. Authorization for access will be determined by checking collection identifiers associated with user identities against a collection database, also maintained by DLPS.

User database

1. The user database will be built from a variety of sources of University of Michigan administrative data and Library user data.
2. The user database will contain all relevant information about the user for the purposes of this application, including:

    a. User name (in discrete fields for last name, first name, etc.)

    b. E-mail address

    c. Institutional affiliation (if applicable)

    d. Status (i.e., faculty, staff, student, external, etc.)

    e. Course enrollment information (if possible, in order to link to University Reserves activities) where applicable

f. Departmental affiliation, where applicable

g.

1. University ID number for UM-affiliated persons
2. a "key" for persons not affiliated with UM

h. A password for all non-UM users and for UM users if requested/generated by the user

i. Collections or collection groups to which the user has access rights.

Collection database

1. The collection database will contain all *discretely accessible* DLPS-managed collections, entered and maintained by DLPS staff.
2. The collection database will contain all relevant information about the collection for the purposes of this application, including:

    a. Collection identifier (locally invented, but perhaps derived from cataloging information in MIRLYN)

    b. Collection name (as cataloged or as supplied by publisher)

    c. Collection source (i.e., frequently publisher, but including any contributor)

    d. Collection description (supplied by staff, preferably selector)

    e. Collection "rights class," where "rights class" is explicated by an access description such as "UM-member" (i.e., all current UM faculty, staff, and students,) but may also be groupings of collections from a particular publisher such as "HRAF-subscriber" (i.e., a user who has subscribed to both the Ethnology and Archaeology collections from HRAF).

    f. Collection "class" (e.g., GUMS, Bib, Imaging)

Authentication

1. The system must be able to support multiple methods of authentication, including Kerberos and passwords specific to the system.

    a. Kerberos identities and passwords will be supported for University of Michigan faculty, staff, and students. The eligibility of the user must first be determined through a lookup in the locally-supported user database. For example, if a user

selects Kerberos as his method of authentication, a call will be made to the user database to determine eligibility, after which the user's password will be checked against the KA Server.

   b. Locally-generated identities and passwords will also be honored. These identity/password pairs will be generated for both UM users and external users. A user will have a "key." This "key" can be used to set or change the user's password.

2. UM persons will use their uniqnames as their user identity. The University ID number that appears on a UM affiliate's ID card will serve as a key. For example, as a UM user, I can use my University ID to assign myself a password to be used in conjunction with my uniqname.

3. External users will be provided a user identity and "key" by the system, under the authorization of the collection provider (e.g., the University of Michigan Press). For example, an unaffiliated user who purchases access to the Middle English Compendium will be provided a "key" by the University of Michigan Press, and with this "key" the user may assign herself a password.

Selective application of encryption

1. The system must use SSL to encrypt authentication and password changes.

2. The system must not use encryption to transfer Library data to the user or to accept user transactions (e.g., collection searching or browsing). (This requirement is solely for performance and reliability but is a firm requirement.)

Authorization

1. The system will use either cookies or annotated URLs to register authorization information with the user's browser.

   a. We recognize that there are critical user and technical issues associated with the use of cookies. However, important library information resources such as JSTOR are turning increasingly to the use of cookies. Consequently, we do not deprecate the use of cookies.

   b. Annotated URLs are, in some cases, susceptible to "theft" by users such that the authorization credentials can be passed to another user during the life of the credentials (see below on authorization duration). However, the nature of the materials and the threat of "theft" with regard to these materials does not lead us to deprecate the use of annotated URLs.

2. Authorization credentials will have a two hour life from the time of last activity, if possible, and otherwise the time of issuance.

3. Explore the use of Webscript (http://www5.oclc.org/downloads/software/webscript/default.htm), used by OCLC and others, and possibly of significant value as a tool in this application.

4. *Ideally,* attempt to ensure that the methods we adopt (e.g., Webscript) are consistent with strategies that use Certificates.

Maintenance

1. Maintenance authorization: The system must support different levels of administrative privilege for maintaining user databases.

    a. The highest level of administrative privilege will be made available to selected DLPS and LSO staff whom we will characterize in this document as "root administrators." Root administrators will be able to view, alter, add, and delete any user record.

    b. A similarly high, but collection-constrained level of administrative privilege will be made available to designated representatives of most collection owners. These persons will be referred to in this document as "collection administrators." Collection administrators will be able to view, alter, add, and delete any user record in those collections managed by the collection administrator.

    c. For the purposes of user support and diagnostics, all DLPS, LSO, and selected Library user support staff will be able to view all user records, regardless of user affiliation or subscription. These persons will be referred to in this document as "read-only users."

2. User record source(s)

    a. All records for UM-affiliated persons will be extracted from administrative databases by LSO staff, and submitted through manual and automated processes to DLPS staff for database updates.

    b. All records for users required for external collection providers (e.g., UM Press) will be created by manual interaction with the system by "collection administrators." Facilities for batch creation of large numbers of user identities will probably be necessary, but are not required for the initial implementation of the system.

3. Administrators (both "root" and "collection") will be able to update the system in real time, with changes being in effect within an hour of update.

User Interface

All interactions between users (i.e., Root administrators, Collection administrators, and read-only users; see Maintenance above) and the database should be possible through standard web browsers, preferably using standard HTML and SSL.

Tool and Resource Selection

This document makes no specific recommendations on tools and resources (e.g., programming languages and hardware). There is some urgency for a better solution, not simply because of the UM user population, but also because of the growing body of external users that DLPS serves. Consequently, we recommend only (and emphatically) that the approach taken use only familiar and tested resources. This project should *not* be used as a platform for becoming familiar with a new tool (e.g., "I'd like to learn more about JAVA, so I'll develop this in JAVA"). After some discussion within DLPS, few specific requirements were articulated beyond a *suggestion* that all parts of the authentication programming be done in the *same* language, and that all components should have source code available to them (e.g., that we would not rely on a compiled object for which we have no source code as a component in the system). Beta versions of software or other tools should be avoided at all costs.

Appendix 1: Context

The problem libraries face, in brief, is that IP addresses alone do not suffice in permitting or restricting access to the valid user or user population. Libraries and publishers have relied primarily on ranges of IP addresses to identify the body of valid users and provide them with access to licensed resources. Increasingly, the body of valid (and active) users accesses the resources from locations where their connection comes from an IP address other than one under the control of the licensing institution (e.g., the University of Michigan). The user may be in a distance learning program, on sabbatical at another institution, outside of the local dialing area, or simply accessing the resource through a third party ISP (e.g., via a cable modem); in all cases, the user works from an IP address outside the range of valid IP addresses. The problem is further complicated by a growing need to serve users who are not members of the typical licensed body of users. For example, most UM resources are licensed to current faculty, students, and staff, but some licenses have been negotiated to include UM Online users (a body of users that consists primarily of alumnae). Finally, we are beginning to face situations where users have rights to access some resources but not others, and so where an IP address alone does not serve to adequately identify the user's identity and his/her rights to specific resources. While this area of the document should perhaps be developed more fully, I hope it will be sufficient to offer only this brief statement of the problem and to point to one of the better statements of the problem, articulated by Cliff Lynch of CNI.

Appendix 2: Miscellaneous Related Issues

We continue to have problems with the current authcookie mechanisms on the Gartner data. All UM Ann Arbor and Dearborn users should be permitted when they authenticate with their Kerberos identities. We know that if a user does not come from a valid umich.edu IP address, the user is not permitted (regardless of authentication by identity). Do we know that Dearborn users have valid Kerberos identities in the umich.edu realm? If not, are we permitting the appropriate realm?

John Price Wilkin
Last revised: 1999.3.17