May 2010

Authentication and Authorization: Security Issues for Institutional Digital Repositories

Md. Zahid Hossain Shoeb
Independent University of Bangladesh, zhshoeb@gmail.com

M Abdus Sobhan
Independent University of Bangladesh, asobhan@secs.iub.edu.bd

Library Philosophy and Practice 2010

ISSN 1522-0222

Authentication and Authorization: Security Issues for Institutional Digital Repositories

Md. Zahid Hossain Shoeb
Independent University, Bangladesh (IUB) Library
Baridhara, Dhaka, Bangladesh

M Abdus Sobhan
School of Engineering & Computer Science
Independent University, Bangladesh (IUB)
Baridhara, Dhaka, Bangladesh

Introduction

In this digital age and the development of Information and Communication Technology (ICT) many organizations have realized the benefits of sharing information within the organization as well within the community and globally. These organizations may be corporate company, research organization or academic institutions. In the academic institutions with the higher education, information capturing, dissemination and sharing is practiced most. In spite of Open Source Drive, in the highly competitive environment, many university or colleges raise a paradox between allowing information and knowledge to flow freely, and the need to keep certain information very secure. In restricted or closed-information environment secured information channel, authorization and authentication of both users and digital contents are a burning issue today. Digital contents are managed and stored in repository to share. Repository of an institution can support research, learning, and administrative processes as well as purposes. Standards are followed for the repositories which ensure that the contents contain is accessible in that and it can be searched and retrieved for later use. A wide variety of contents may be included in the digital repositories for the multiplicity of purposes and users. It is the technical ability and administrative policy decision that what kind of materials goes into a repository (Jones, et al 2006). A proper digital repository not only requires an organized collection of digitized content, it also requires that the content be accessed and distributed as widely as possible to legitimate users around the globe. Access management and control is one of the major concerns for content-providers on the Internet. Without a proper access management mechanism confidentiality and integrity of information cannot be guaranteed. Different conventional methods are practiced by the content-providers but not a single method is sufficient for access management (Ray and Chakraborty, 2006). However, the administrators of the digital content-providers mostly expect their preferences for the technology or the procedure to be available which may be best practiced globally.

Methodology

Initially, substantial amount of literatures have been reviewed to come up with an idea for formulating this paper which is a review by type. The researches, practices, progresses, development and successes for the access management specially authentication and authorization are reviewed to see the global practices by the repository administrators or managers. Even in Bangladesh there are very few repositories, and the repository managers are interviewed by the author though all are examined also to see the status. Though there are very common types of process or methods observed where traditional or

built in securities of repository software or operating systems are adopted most. Based on review and local managers interview this paper gives idea about the current practices about authentication and authorization.

Access Management

Access management typically is a combination of users' authentication and authorization, access permission operations, policies for license agreement and digital materials authentications or digital rights management. Authentication is the process of determining the validity of a user who claims to be, and authorization is the process of determining what resources a user is permitted to access. Digital Rights Management (DRM) is a system of solutions created or designed as a means to prevent unauthorized access, duplication and illegal distribution of copyrighted digital media. In online environment, the scope of DRM can be leveraged to control access to and usage of digital objects and to impose restrictions on their misuse (Functional Groups, 2009). Access Management ensures security of resources on servers but also during communication to ensure authenticity and integrity of data. It is possible for an unauthorized user to snoop on communication between a user's browser and a Web server and hack sensitive information. Occurrences of unauthorized user getting access to important Web sites and defacing them are not uncommon. Electronic content can be copied very easily, it is essential to impose measures to control misuse of digital content. IP authentication and password-based access, two most commonly used authentication methods, are not able to protect the content from being duplicated or shared. Access Management is necessary most for commercial digital contents because their access is restricted to its subscribers or licensed users. Even when access to digital collections is provided openly, access control is required for assigning responsibilities for operations such as, additions, updating, editing and deleting or with-drawing content, and other tasks related to digital collections. Other reasons to control access to materials in a repository may include confidentiality of resources. Tracking of all changes made so that the collections can be restored if any system error occurred. In access management as other matters are related to policy or administrative decisions the user authentication, authorization and digital material authentication are most necessary issues.

User Authentication: User Validity

A user authenticates with his or her organizational or personal identification. The identity provider passes the minimal identity information necessary to the service manager for authentication to enable an authentication decision (ACM, 2009). Digital identities are increasingly being used to facilitate the execution of transactions in various domains. When developing and analyzing digital identity technologies, it is important to consider the type and objective of repository, type of digital content, security of the system, security of communication channel, diversity of users' platform, number of users, even the perceptions and responses of end users also. Different authentication processes are as follows:

The most common and familiar authentication process is Log-in ID and Password-based Access (Antón, 2007). Log-in is also called log on, sign in, or sign on which identifies oneself to the system in order to obtain access. The primary use of a computer login procedure is to authenticate the identity of any computer user or computer software attempting to access the computer's services (Logging, 2002). Another popular authentication process is IP Filtering or IP authentication. This process is a packet filter that analyzes TCP/IP packets. That is software routine that analyzes incoming data packets and forwards them or discards them based on one or more criteria such as address, range of addresses and types (IP filter, 2009). Institutions or organizations are encouraged to register for accessing digital contents using IP addresses (ranges) if they are static. This allows for: seamless access (no logon screen), usage statistics for the institution, greater security as no misuse of usernames and passwords, can allow for access for all computers on campus to resources and much more (INASP, 2009). Web Cookie is another process of user authentication, which can be used by a server to recognize previously-authenticated users and to personalize the web pages of a site depending on the preferences of a user (Cookies, 2009). A cookie is a token that the web browser stores on disk in the form of a small text file. Cookies provide a way to track individual users' usage of website. Web Proxy is another way to authenticate. Most proxy programs like

Squid, NetCache provide a mechanism to deny access to certain URLs in a blacklist, thus providing content filtering. A content filtering proxy will often support user authentication, to control web access. EZproxy (2008) is also a web proxy server program that provides users with remote access to Web-based licensed content offered by libraries. It is middleware that authenticates library users against local authentication systems and provides remote access to licensed content based on the user's authorization (Proxy server, 2008). Another method Challenge-Response Authentication is used to prove the identity of a user logging into the network. When a user logs on, the network access server, wireless access point or authentication server creates a “challenge,” which is typically a random number sent to the client machine. The client software uses its password or a secret key to encrypt the challenge via an encryption algorithm or a one-way hash function and sends the result back to the network (the “response”). The authentication system also performs the same cryptographic process on the challenge and compares its result to the response from the client. If they match, the authentication system has verified that the user has the correct password (Challenge-response authentication, 2008). Referring URL is a way to authenticate users. From the point of view of a web page or resource, the referrer, or HTTP referrer, identifies the address of the webpage or URL, the more generic URI of the resource which links to it. By checking the referrer, the new page can see where the request came from. Referrer logging is used to allow websites and web servers to identify where people are visiting them from, for promotional or security purposes. Referrer is a popular tool to combat Cross-site request forgery, but such security mechanisms do not work when the referrer is disabled (HTTP Referrer, 2008). Referrer is widely used for statistical purposes.

User Authorization: Resource Access Permission

Authorization defines users' permissions in terms of access to digital resources and extent of its usage. Authorization is granted to the successfully authenticate users according to his/her rights information available in the Access Management System (AMS) (Lynch, 2009). Authorization also addresses the issue of responsibilities assigned to different personnel involved in development of a digital repository/library and their respective authorities in terms of addition, deletion, editing and uploading of records into a digital collection. Authorization is more challenging than authentication, especially for widely distributed digital content providers.

Conventional access control architecture denotes an access control policy as a subject (user) is authorized to exercise some permission on an object. This usual model implicitly assumes that the user population is known more or less. But in a digital content environment the user population is vast, dynamic and impossible to predict all the users. Thus conventional authorization or access control mechanisms that rely on knowing the user and associating permissions with them fail significantly in digital repositories. So, this digital environment demands some further challenge for access control (Bertino, 2002). The access control policies are often based on user qualifications and characteristics. In one of the early works on access control in digital repositories or libraries, Gladney (1997) proposes a scheme called DACM (Document Access Control Methods), where the basic idea is geared toward flexible access control with some extensions to handle mandatory access control. Blaze has proposed credential-based access control (Blaze, 1996), to address the problem of unknown users. In these models a user has to produce one or more testimonials that have been certified by one or more third parties. The credential provides information about the rights, qualifications, responsibilities and other characteristics attributable to its bearer by the third parties. These third parties need to be trusted by the service provider. Winslett (1997) developed a credential based security and privacy related system for enforcing access control in digital contents of repository or system. Access to systems containing protected information resources must be managed based on one or multiple selections of the alternative access control methods. However, different methods are based on, Users identity, Role, Policy, Content Dependency, Context, View, Time, Physical Location, Network Node, Mandatory, and Discretionary. In-addition, a risk assessment is needed to conduct to identify the data or resource risk and severity prior to establishing the level and selection of access controls or authorization to digital contents (Access control, 2009).

Digital Materials Authentication

Digital Watermarking and Digital Signature are very common in use to provide a range of solutions for identifying, securing, managing and tracking digital materials (Stallings, 2003). In different types of objects like audio, video, still images and printed documents *Digital Watermarking* technologies allow users to embed. This technology permits digital code that is imperceptible during normal use but readable by computers and software. The major purpose of digital watermarks is to provide protection for intellectual property that is in digital format. This system does not prevent copying, but ensures that any copies made of the media will be traceable to a particular copy and perhaps to a particular user. In this process, also referred to as data embedding, information hiding, or simply watermarking, a pattern of bits is inserted into a digital image, audio or video file that identifies the file's ownership and can convey additional information like copyright. Unlike printed watermarks, which are intended to be somewhat visible, digital watermarks are designed to be completely invisible, or in the case of audio clips, inaudible. Moreover, the actual bits representing the watermark must be scattered throughout the file in such a way that they cannot be identified and manipulated. And finally, the digital watermark must be robust enough so that it can withstand normal changes to the file, such as rotation, filtering or the application of compression algorithms such as JPEG that discard some of the original data (Watermarking, 2009). On the other hand *Digital Signature* is an electronic signature which authenticates to identify the sender of the message where original document has been sent remain unchanged. It is easily transportable, cannot be reproduce by someone else, and can be automatically time-stamped. In addition the message is sent, the sender cannot easily reject it later. A digital signature can be used with any kind of message, whether it is encrypted or not, simply so that the receiver can be sure of the sender's identity and that the message arrived intact. A digital certificate contains the digital signature of the certificate-issuing authority so that anyone can verify that the certificate is real (Digital Signature, 2009).

A digital signature scheme typically consists of three algorithms:

a. A key generation algorithm that selects a private key uniformly at random from a set of possible private keys. The algorithm outputs the private key and a corresponding public key.

b. A signing algorithm which, given a message and a private key, produces a signature.

c. A signature verifying algorithm which given a message, public key and a signature, either accepts or rejects.

Two main properties are required. First, a signature generated from a fixed message and fixed private key should verify on that message and the corresponding public key. Secondly, it should be computationally infeasible to generate a valid signature for a party who does not possess the private key (Brands, 2000) .

Conclusion

As institutions, information strategies which call for sharing and licensing access to information resources in the networked environment, authentication and access management have emerged as major issues which threaten to hinder progress. While considerable work has been done over the last two decades on authentication within institutions and, more recently, in support of digital repository, a series of new technical and policy issues emerge in the cross-organizational authentication and access management context. In this paper it has been illustrated the secured process digital repository which ultimate objective is information sharing and dissemination. Though it has not shown any model or architecture regarding any types of technical aspects whether it has been tried to introduce the terms and methods or even the factors for access management for digital content provider. A lot of work, however still remains to be done. In future the authors will work on few open source technology for access

management of digital library by which the digital content provider can find a solution to gather the working process of those for implementation.

References

About Digital Watermarking. Available: http://www.willamette.edu/wits/idc/mmcamp/watermarking.htm

Access control criteria for right to use automated information resources. Available: http://michigan.gov/documents/Policy_1350_157471_7.40_Access_Control_Final_PDF.pdf

Antón, L., Jones, A., Earp, J.B. (2007). Towards understanding user perceptions of authenticationtechnologies. *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society.* Virginia : ACM.

Bertino, E., Ferrari, E., & Perego, E. (2002). Max: An access control system for digital libraries and the web. Oxford, UK: *Proceedings of the 26th IEEE International Computer Software and Applications Conference.*

Blaze, M., Feigenbaum, J., & Lacy, J. (1996). Decentralized trust management. Oakland, CA: *Proceedings of the 1996 IEEE Symposium on Security and Privacy.*

Brands, S.A. (2000). *Rethinking public key infrastructures and digital certificates: Building in privacy*. London: MIT Press.

Challenge-response authentication definition. Available: http://encyclopedia2.thefreedictionary.com/Challenge-response+authentication

EZproxy: OCLC – Web and Data Services. Available: http://www.oclc.org/us/en/ezproxy/default.htm

FAQs on Information resources. Available: http://www.inasp.info/file/188/faqs-on-information-resources.html

Functional groups: Access management R & D. Available: http://www.inflibnet.ac.in/functionalgroup/openaccess.html

Gladney, H.M. (1997). Access control for large collections. *ACM Transactions on Information Systems 15:* 154–194.

HTTP referrer. http://en.wikipedia.org/wiki/Referrer

IP filter definition of IP filter in the free online encyclopedia. Available: ttp://encyclopedia2.thefreedictionary.com/IP+filter

Jones, R., et al. (2006). *The Institutional Repository*. Oxford: Chandos.

Logging (computer security). Available: http://en.wikipedia.org/wiki/Logging_%28computer_security%29

Lynch, C. (n.d.). A white paper on authentication and access management issues in cross-organizational use of networked information resources. http://www.cni.org/projects/authentication/CNI_authentication.doc

Proxy server. http://en.wikipedia.org/wiki/Web_proxy

Ray, I., & Chakraborty. S. (2006). A framework for flexible access control in digital library systems. *Data and Applications Security* : 252–266.

Stallings, W. (2003). *Cryptography and network security: Principles and practice* . New Delhi: Pearson Education.

Towards understanding user perceptions of authentication technologies. http://portal.acm.org/citation.cfm?doid=1314333.1314352

Web handbook – Cookies. Available: http://archive.cabinetoffice.gov.uk/e-government/resources/handbook/html/4-7.asp

Winslett, M., et al. (1997). Assuring security and privacy for digital library transactions on the Web: Client and server security policies. *Proceedings of the IEEE international forum on Research and Technology Advances in Digital Libraries,* Washington , DC , USA , pp. 140-151.

What is digital signature? – a definition from whatis.com. Available: http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci211953,00.html