DESIDOC Journal of Library & Information Technology, Vol. 29, No. 4, July 2009, pp. 21-27

Access Management for Digital Repository

Md. Zahid Hossain Shoeb

Independent University, Bangladesh
58, Park Road, Baridhara, Dhaka 1212, Bangladesh
E-mail: zahid@iub.edu.bd

ABSTRACT

Access management illustrates the process of providing controlled, and secure access to resources. It involves both authentication—the process of determining the validity of a user who claims to be, and autorisation—the process of determining what resources a user is permitted to access. Access management is significant as an important as well as one of the key services for the network infrastructure of digital repository activities for a wide range of audiences. Like the digital collection of digital repository, it also requires that the content be accessed and distributed as widely as possible to valid users around the globe. Access management and control is one of the major concerns for digital content authority over the Internet. Confidentiality and integrity of information cannot be ensured without a proper access management mechanism. Though most of the Institutional Repositories or Open Access Repositories are satisfied with the current mechanisms of the Open Source Software, which they use to build the repositories, restricted repositories or digital libraries demand more security and authenticity. The paper discusses the current practices and issues of access management for digital repositories including user authentications, user authorisation, authentication, and the technology of secured digital communication of digital materials. This paper also gives a basic idea about access management practices in Bangladesh by the digital content providers.

Keywords: Access control, authorisation, digital communication, digital repository, security, Bangladesh

1. INTRODUCTION

A digital repository is a mechanism for managing and storing trustworthy digital contents. By focus, repositories can be subject, institutional or commercial. Repository of an institution can support research, learning, and administrative processes as well as purposes. Standards are followed by the repositories, which ensure that its contents are accessible, and can be searched and retrieved for use at a later stage. A wide variety of contents may be included in digital repositories for the multiplicity of purposes, and users. It is the technical ability and the administrative policy which decides what kind of materials goes into a repository. Usually, contents can include research outputs such as journal articles or research data, e-theses, e-learning objects and teaching materials, and administrative data.

Some repositories only take in particular items like theses or journal papers even as others gather any reliable scholarly work produced by the institution. Complex objects and other data files are also a part of digital contents[1]. A true digital repository not only requires an organised collection of digitised 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[2].

Without a proper access management mechanism confidentiality and integrity of information cannot be guaranteed. Though different methods are practiced by the content providers, not a single method is sufficient for access management. However, the administrators of the digital content providers prefer the best technology or the procedures available globally, and always try to do some innovative reliable jobs to provide better accessibility to the users. In this paper, the author does not intent to illustrate the underlying complexities, or defining any technical architecture of the methods or systems, but has attempted to explore the current practices being followed for the digital repository.

The paper discusses the access management, different types of user authentication, users authorisation,

Received on 2 February 2009

authorisation of digital materials in digital repositories and the current scenario of digital repositories in Bangladesh.

2. ACCESS MANAGEMENT

Access management is typically a combination of users' authentication and authorisation, access permission operations, policies for license agreement, and digital materials authentications, or digital rights management (DRM). DRM is a system of solutions created or designed as a means to prevent unauthorised 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[3].

Access management not only ensures security of resources on servers but also during communication to ensure authenticity and integrity of data. It is possible for an unauthorised user to snoop on communication between a user's browser and a web server and hack sensitive information. Occurrences of unauthorised user getting access to important websites and defacing them are not uncommon. Techniques of data encryption are used widely for communicating sensitive information such as user's password and personal identification number (PIN). Encryption renders data unintelligible and unusable even if accessed by an unauthorised person.

Electronic content can be copied very easily, it is essential to impose measures to control misuse of digital content. Password-based access and IP authentication, two most commonly used authentication methods, are not enough to protect the contents from being duplicated or shared. Access management is necessary 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, updation, editing and deleting or withdrawing content, and other tasks related to digital collections. Other reasons to control access to material in a repository may include confidentiality of resources. Tracking of all changes is made so that the collections can be restored if any system error occurred. The user authentication, authorisation, and digital material authentication are necessary in access management as other matters are related to policy or administrative decisions.

3. USER AUTHENTICATION

A user authenticates with his or her organisational or personal identification. The identity provider passes the minimal identity information necessary to the service manager for authentication to enable an authentication decision[4]. Digital identities are increasingly being used to facilitate the execution of transactions in various domains. When developing and analysing 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. The following techniques are commonly used for user authentication.

3.1 Login ID and Password-based Access

The most common and familiar authentication process is login ID and password-based access[5]. Log-in is also called logon, sign-in, or sign-on to 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 attempting to access the computer's services. To login to a system usually requires a user name—a unique sequence of characters the user chooses to represent him/herself. Many websites now use e-mails as username. A password, another sequence of characters, provides the user with a key to the system and is kept secret from others[6].

3.2 IP Filtering or IP Authentication

This process is a packet filter that analyses TCP/IP packets. It is a software routine to analyse and forward/discard incoming data packets based on one or more criteria such as address, range of addresses, and types[7]. Institutions or organisations are encouraged to register for accessing digital contents using IP addresses (ranges) if they are static. This allows (i) seamless access (no logon screen), (ii) usage statistics for the institution, (iii) greater security as there is no misuse of usernames and passwords, (iv) access to all computers thereby releasing other terminals and staff time, and (v) direct recognition of institutional networks by publishers and vice versa[8].

3.3 Web Cookies

A cookie is a token that the web browser stores on disk in the form of a small text file. Cookies can be used by a server to recognise previously-authenticated users and to personalise the web pages of a site depending on the preferences of a user. Cookies provide a way to track individual users' usage of website[9].

3.4 Web Proxy

The most common use of a web proxy is to serve as a web cache. 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[10]. EZproxy 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 contents based on the user's authorisation[11].

3.5 Challenge-Response Authentication

This method 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 using 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 with the response from the client. If they match, the authentication system has verified that the user has the correct password[12].

3.6 Referring URL

From the point of view of a web page or resource, the referer (HTTP referer), identifies the address of the web page or URL, the more generic URI of the resource which links to it. By checking the referer, the new page can see where the request came from. Referer logging is used to allow websites and web servers to identify where people are visiting from for promotional or security purposes. Though Referer is a popular tool to combat cross-site request forgery, such security mechanisms do not work when the referer is disabled. Referer is widely used for statistical purposes[13].

3.7 Biometric Technologies

Biometric is an automated method of recognising a person based on a physiological or behavioural characteristic. Among the features measured are: face, fingerprints, hand geometry, handwriting, iris, retinal, vein, and voice. Biometric technologies are becoming the foundation of an extensive array of highly secure identification and personal verification solutions. As the level of security breaches and transaction fraud increases, the need for highly secure identification and personal verification technologies is becoming apparent[14].

4. USER AUTHORISATION

While the process of authentication ascertains the identity of a user, authorisation defines his or her permission to access e-resources and extent of its usage. Authorisation is granted to the successfully authenticate users according to his/her rights information available in the Access Management System (AMS). Authorisation also addresses the issue of responsibilities assigned to different persons 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. Authorisation is more challenging than authentication, especially for widely distributed digital content providers.

Traditional access control architecture denotes an access control policy as a subject (user) and is authorised to exercise some permissions on an object. This conventional model implicitly assumes that the user population is known more or less. But in a digital content environment, the user population is vast and dynamic. Thus, conventional authorisation or ACM that rely on knowing the user and associating permissions with them fail significantly in digital repositories. So, digital environment demands further challenges for the access control[15].

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[16] proposed 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 to address the problem of unknown users[17]. In these models, a user has to produce one or more testimonials 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[18] 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 ACMs. The conventional methods are:

- UBAC (User-based Access Control) also called identity-based access control for defining permissions for each user.
- RBAC (Role-based Access Control) uses a profile to map a group of users to resources.
- PBAC (Policy-based Access Control) is a set of rules that determine access rights.
- CDAC (Content Dependent Access Control) limits the rights of users based on content of resource to specific fields (range) or cells (data point) typically within a database.
- CBAC (Context-based Access Control) includes sequence of events that preceded the access

attempt to grant or deny access.

- VBAC (View-based Access Control) uses pre-defined interfaces (views) so that users only gain access to sub-resources within allowed view.
- TBAC (Time-based Access Control) limits access to resources based on time associations.
- PLAC (Physical Location Access Control) limits access to resources by a given location.
- NNAC (Network Node Access Control) limits access to specific nodes or networks.
- MAC (Mandatory Access Controls) are required controls.
- DAC (Discretionary Access Controls) are optional to the user.

In addition, a risk assessment is needed to identify the data or resource risk, and severity prior to establishing the level and selection of access controls or authorisation to digital contents[19].

5. AUTHENTICATION OF DIGITAL MATERIALS

To provide a range of solutions for identifying, securing, managing and tracking digital materials, Digital Watermarking and Digital Signature are very common in use.

5.1 Digital Watermarking

Digital watermarking technologies allow embedding of digital codes into different types of objects like audio, video, still images, and printed documents. Digital codes are 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. The name comes from the faintly visible watermarks imprinted on stationery that identify the manufacturer of the stationery.

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. Digital watermark must be robust enough to 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 (lossy compression)[20].

5.2 Digital Signature

A digital signature is an electronic signature that can be used to authenticate the identity of the sender of a message or the signer of a document possibly to ensure that the original content of the message or document that has been sent is unchanged. Digital signatures are easily transportable, cannot be reproduced by someone else, and can be automatically time-stamped. The ability to ensure that the original signed message arrived means that 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[21].

A digital signature scheme typically consists of the following three algorithms:

- 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.
- A signing algorithm, which given a message and a private key, produces a signature.
- A signature verifying algorithm, which given a message, public key and a signature, either accepts or rejects.

A signature generated from a fixed message and fixed private key should verify that message and the corresponding public key. It should also be computationally infeasible to generate a valid signature for a party who does not possess the private key[22].

6. TECHNOLOGY OF SECURED DIGITAL COMMUNICATION

For fast, reliable, authenticate access to digital contents a secure communication channel is required. Few secure technologies are:

6.1 Cryptography and Encryption

In cryptography, encryption is the process of converting information (plain text or numbers) from its normal, comprehensible form into an incomprehensible encrypted format rendering it unreadable except for those

who possess special knowledge, usually referred to as a key[23]. Encryption is used in DRM to restrict the use of copyrighted material, and in software, protection against reverse engineering and software piracy. Standards and cryptographic software and hardware to perform encryption are widely available. Software used for encryption can also be used to perform decryption, i.e., to make the encrypted information readable again.

6.2 Digital Certificates

Digital certificates are electronic files used to authenticate web resources, users, and organisations over the Internet to ensure integrity of content. Digital certificates are part of a technology called Public Key Infrastructure (PKI) that includes Certification Authorities (CAs) such as GeoTrust, VeriSign, Commodo, etc. CA issues, manages, and revokes digital certificates to relying parties who use the certificates as indicators of authentication, and clients who request, manage, and use certificates[24].

Organisations that use digital certificates to authenticate their users, maintain a database or directory, using a directory access protocol called LDAP that stores information about certificate holders and their certificates. Digital certificates are based on public-key cryptography that uses a pair of keys (private and public) for encryption and decryption. It contains the name, a serial number, expiry dates, copy of the certificate holder's public key, and the digital signature of the CA, so that a recipient can verify that the certificate is genuine. These electronic credentials assure that the keys actually belong to the person or the specified organisation. Messages can be encrypted with either the public or the private key and then decrypted with the other key. The recipient of an encrypted message uses the CA's public key to decode the digital certificate attached to the message, and then obtains the sender's public key and identification information held within the certificate. With this information, the recipient can send an encrypted reply. Digital certificates form the basis for secure communication, and client and server authentication on the web.

6.3 Transport Layer Security

Transport Layer Security (TLS)[25] and its earlier version Secure Sockets Layer (SSL), are cryptographic protocols that provide authentication and communication privacy to ensure secure communication on the Internet for applications such as web browsing, e-mail, Internet faxing, and instant messaging. The protocols involve, between client and server, exchange of public key and digital certificates-based authentication, and symmetric encryption of data. The exchange process initialises when a client connects to a TLS-enabled server requesting for a secure connection, and presents a list of supported encryption codes (ciphers and hashes). The server agrees on one of such encryption codes that it also supports and notifies the client of the decision. The server sends back its identification in the form of a digital certificate.

6.4 Simple Authentication and Security Layer

Simple authentication and security layer **(**SASL) is a framework for authentication and data security in Internet protocols. It decouples authentication mechanisms from application protocols allowing any authentication mechanism supported by SASL to be used in any application protocol that uses SASL[26]. Authentication mechanisms can also provide a data security layer offering data integrity and data confidentiality services. Application protocols that support SASL typically also support TLS. SASL is an Internet Engineering Task Force (IETF) protocol, Protocols currently supporting SASL include Internet Message Access Protocol (IMAP), Lightweight Directory Access Protocol (LDAP), Post Office Protocol (POP), and Simple Mail Transfer Protocol (SMTP).

6.5 X.509

X.509 is a technical standard of the International Telecommunication Union for the PKI[27]. X.509 specifies, amongst other things, standard formats for public key certificates and a certification path validation algorithm. Under the system, a CA issues a certificate along with a public key to a particular name, or to an alternative name such as an e-mail address or a DNS-entry. An organisation's trusted root certificates can be distributed to all employees so that they can use the company PKI. It is left to the browsers' owners to determine which CAs to trust. Moreover, it is customisable.

6.6 Security Assertion Markup Language

Security Assertion Markup Language (SAML) is an XML standard for exchanging authentication and authorisation of data between security domains, i.e. between an identity provider (a producer of assertions) and a service provider (a consumer of assertions). SAML assumes that a user has enrolled with at least one identity provider[28]. This identity provider is expected to provide local authentication services to the user. At the user's request, the identity provider passes a SAML assertion to the service provider. On the basis of this assertion, the service provider makes an access control decision.

6.7 Open Digital Rights Language

Open Digital Rights Language (ODRL) is an XML-based standard Rights Expression Language (REL) used in digital rights and open content management systems[29]. ODRL is managed by an open organisation open to public participation. The Open Mobile Alliance has adopted ODRL, and new mobile phone handsets support this ODRL Profile.

6.8 eXtensible Rights Markup Language

eXtensible Rights Markup Language (Xrml) is also a REL owned by ContentGuard. It is used to control the content with the help of licensing policies. XrML is based on XML and describes rights, fees and conditions together with message integrity, and entity authentication information. XrML provides a universal method for securely specifying and managing rights and conditions associated with all kinds of resources including digital content as well as services. XrML 2.0 is extensible and compliant with XML namespaces using XML schema technology.

7. ACCESS MANAGEMENT PRACTICES IN BANGLADESH

Different types of digital repositories contain a wide variety of objects which need security, integrity, authentication, and authorisation[30]. Most of the technologies used for making, preserving or distributing digital contents, contain built-in security mechanisms. Sometimes, these are not sufficient for managing different types of users with different interest and age. So, customisation, third party technology, or more trusted mechanisms are required followed by a combination of different technology for better performance and for suitable solution. As a developing country, Bangladesh is a land of dense population, and is growing fast in educational and non-governmental sectors. Unfortunately in Bangladesh, other than Independent University, Bangladesh (IUB), International Centre for Diarroeal Disease Research (ICDDR), and BRAC University, not many digital repository has been built. ICDDR and recently BRAC University have established and hosting institutional repository using DSpace the minimum standard by functionality, usability and architecture. Both ICDDR and BRAC university host their research and scholarly output in their repository in small scale. The digital library in School of Engineering and Computer Science (SECS) of IUB is not a mature effort though the initiative is good.

8. CONCLUSION

Access management is important for online information management. Four aspects of access management are users' authentication and authorisation, access permission operations, policies for license agreement, and digital materials authentications or DRM. The paper discussed the practices of access management, especially authentication, and access control of digital repository. More awareness is required for using these repositories but unfortunately the repositories are not focusing the users in Bangladesh, and using login id and password-based access. However, in this day of digital environment, more open access digital repositories are needed for dissemination of digital contents where there will be no physical boundary, and round-the-clock availability of the resources and multifaceted objects.

REFERENCES

1. Jones, R. *et al.* The Institutional Repository. Chandos, Oxford, 2006.

2. Ray, I. & Chakraborty, S.A. Framework for flexible access control in digital library systems. *In* Data and Applications Security, 2006, pp. 252–66.

3. Functional groups: Access management R & D. http://www.inflibnet.ac.in/functionalgroup/openaccess. html (accessed in June 2008).

4. Towards understanding user perceptions of authentication technologies. http://portal.acm.org/citation.cfm?doid=1314333.1314352 (accessed in May 2008).

5. Anton, A.I.; Jones, L.A. & Earp, J.B.. Towards understanding user perceptions of authentication technologies. *In* Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society, Virginia, USA. 2007.

6. Logging (computer security). http://en.wikipedia.org/wiki/Logging%28computer_ security%29 (accessed in September 2008).

7. IP filter definition of IP filter in the free online encyclopedia. http://encyclopedia2.thefreedictionary.com/IP+filter (accessed in March 2008).

8. FAQs on information resources. http://www.inasp.info/file/188/faqs-on-information-resources.html. (accessed in May 2008).

9. Web Handbook: Cookies. http://archive.cabinetoffice.gov.uk/e-government/resources/handbook/html/4-7.asp (accessed in September 2008).

10. EZproxy: OCLC–web and data services. http://www.oclc.org/us/en/ezproxy/default.htm (accessed in August 2008).

11. Proxy server. http://en.wikipedia.org/wiki/Web_proxy. (accessed in August 2008).

12. Challenge-response authentication definition. http://encyclopedia2.thefreedictionary.com/Challenge-response+authentication (accessed in September 2008).

13. http://en.wikipedia.org/wiki/Referrer (accessed in September 2008).

14. Biometric technology: AIM global. http://www.aimglobal.org/technologies/biometrics (accessed in October 2008).

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

16. Gladney, H.M. Access control for large collections. *ACM Trans. Inform. Sys.,* 1997, **15**, 154–94, .

17. Blaze, M.; Feigenbaum, J. & Lacy, J. Decentralised trust management. *In* Proceedings of the 1996 IEEE Symposium on Security and Privacy. Oakland, CA, 1996.

18. Winslett, M. *et al*. Assuring security and privacy for digital library transactions on the web: Client and server security policies. *In* Proceedings of the IEEE international forum on Research and Technology Advances in Digital Libraries, Washington, DC, USA. 1997, pp. 140-51,.

19. Access control criteria for right to use automated information resources. http://michigan.gov/documents/Policy_1350_157471_7.40_Access_Control_ Final_ PDF. pdf (accessed in September 2008).

20. About digital watermarking. http://www.willamette.edu/wits/idc/mmcamp/watermarking.htm (accessed in September 2008).

21. What is digital signature? http://searchsecurity.techtarget.com/sDefinition/0,sid14_ gci2 11953,00.html (accessed in September 2008).

22. Digital_signature. http://en.wikipedia.org/wiki/Digital_signature (accessed in September 2008).

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

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

25. Transport layer security. http://www.mnet.state.mn.us/mail/setup/tls.php (accessed in December 2008).

26. Simple authentication and security layer. http://asg.web.cmu.edu/sasl (accessed in December 2008).

27. X.509. http://www.networksorcery.com/enp/data/x509.htm (accessed in December 2008).

28. Security assertion markup language. http://www.service-architecture.com/web-services/articles/security_assertion_markup_language_saml

29. Open digital rights language Version 1.1. http://www.w3.org/TR/odrl (accessed in December 2008).

30. XrML—The digital rights language for trusted content and services. http://www.xrml.org/about.asp (accessed in December 2008).