Yadigar N. Imamverdiyev
Institute of Information Technology of ANAS, Baku, Azerbaijan
yadigar@lan.ab.az

DOI: 10.25045/jpis.v08.i2.08

IDEA4SP: AN INFORMATION SECURITY MODEL FOR DIGITAL LIBRARIES

Digital libraries are complex information systems, which provide online library and information services, and ensuring information security and personal data protection are important for such systems. This paper proposes IDEA4SP-model for the analysis of information security problems in digital libraries, and structures the actual problems by the levels of this model. Moreover, it analyzes the technological approaches to ensuring personal data security and intellectual property protection in digital libraries.

Keywords: *digital library; personal data; information security; privacy; IDEA4SP.*

Introduction

Digital libraries are the integration of complex computer and information systems, and information security is one of the most important problems in their design, construction and operation [1, 2].

The facts that digital libraries are the subject of large-scale cyber attacks, still remain unknown to the general public. Some authors believe that the reason for this is that the number of users using digital libraries is considerably lower than the number of users using online payment systems, social networks and other popular web resources; Library users do not typically have bank information (credit card), which is the target of malware; It is not so valuable and attractive for hackers what people read [2, 3].

However, digital libraries with the Internet connection have always had many security threats. Viruses, Trojans, and other malicious software can destroy online data and documents, steal the data of the organizations and individuals. Unauthorized users can violate the data confidentiality or comprehensiveness through the cyber attacks and security gaps in digital libraries. This, in turn, can weaken the credibility of publishers and content providers, damaging the reputation of the digital library owners, and even causing economic losses, which can lead to anxiety and other concerns when the urgently needed information is not accessible [4].

Due to the varied diversity of actors working with digital libraries, many security requirements are to be taken into account. Each of these actors has different security requirements [4, 5]. For example, a digital library content provider may be interested in protecting the intellectual property rights and the use of content, while a digital library user concerns about the access to content stored in the library. Security requirements based on these needs sometimes contradict and even further complicate the security architecture of those digital libraries.

In Azerbaijan, a number of measures have been taken in the field of creation of digital libraries during the years of independence. Digital libraries equipped with the advanced technical equipments and integrated into the international systems have been developed. Data security of digital libraries, which is an important aspect of the national information space, is of great significance [6-11].

The article aims at analyzing the status of the scientific and practical research on the data security and confidentiality in the digital libraries and identifying the actual research trends.

Digital libraries: content and services

The field of activity related to digital libraries is relatively new, and therefore, there is no fixed terminology yet. The question of "digitization" of libraries was thoroughly reviewed by F.Lancaster in early 1980s [12]. However, his works were futuristic to some extent.

Significant improvements in digital libraries were achieved in the 1990s. Then, the appropriate computing tools and information technologies were developed to ensure the safe

storage, operational processing and effective utilization of various types of information, especially text information. During that period, some countries have started to launch digital library projects.

The concept of "electronic library" began to be concretized, and its goals, objectives and functions were clarified in the course of these works. However, this did not lead to a single-dimensional writing of this problem. The concept of "digital library" also has synonyms as "electronic library" and "virtual library".

The reasonable definition based on the analysis of the practice in the subject field is as follows [13]: "E-library is an information system that enables to efficiently store and use a set of various electronic documents (text, images, audio, video, etc.) located within the system or accessible through the telecommunications network.

The integration of the information resources is intended to be combined with the use of different types of information, its properties and characteristics of the user. This does not mean the absolute physical unity of resources. It may be virtual. The main point is to ensure that the user understands the available information as a single information space.

The efficient navigation is an opportunity for the user to find out exactly what he/she is interested in with the least effort.

Due to the integration of various information resources and effective navigation capabilities, digital libraries are becoming an online system that offers access to a wide range of content and services.

The digital library contents include data and metadata describing various aspects of data (description type, creator, owner, reproduction rights, etc.) and the references to other data/(metadata) and their relationships. Digital library content covers the library catalogue, digital collection, subscription database, electronic magazines, electronic orders, and delivery of documents.

Digital library services include *Open Public Access Catalog (OPAC)*, meta-search, personal portal, book (article) order, guide, and resource search.

Metadata is used to describe the intelligent and technical attributes of the resource objects. Many digital objects are delivered to the user directly through the web; however, some of them may require special application software. The storage can be distributed worldwide through a subscription and collaboration, not only within an organization.

IDEA4SP model for information security

Recently, researchers have proposed numerous digital library models [14-19]. Cornell University researchers have proposed a digital library architecture called *Dienst* [16]. This architecture provides a protocol and the realization of the Internet access to the non-centralized and distributed collection. European researchers proposed a *DELOS* etalon model to build digital libraries [17]. The *DELOS* etalon model is based on 6 basic concepts of digital library building, which are content, user, functionality, architecture, quality and politics. Other models also have different views on this issue [18,19]. The *5S* platform reviews the specification of digital libraries as a definition of the *5S* construction element. A brief description of each *'S'* is provided below [19].

Streams - identify different types and formats of the content stored in digital libraries. The content may be static (such as text data) or dynamic (e.g., *GPS* coordinates of the moving object).

Structures - contain the structure of a digital object (as a whole or its parts). For example, the book can be structured into sections, chapters, subsections, and paragraphs.

Spaces – describe a digital library user interface and index search model. It may be defined as a set of objects and operations that are subject to certain restrictions.

Scenarios – cover the functions, operations, requirements, services and implementations. One of the important scenarios is to describe all possible ways of using the system to perform a user function. Scenarios describe what is going on in the streams in the spaces through the structures, and who is responsible for those operations in the society.

Societies - defined as the "sets of subjects and relationships among them". The subjects envisage the people, hardware and software components.

In terms of information security, three main goals of any library can be defined: 1) confidentiality - digital library contents are not accessible to the unauthorized user; 2) completeness - information is available without any distortion; it indicates that a digital object (resource) can not be changed by an unauthorized person; 3) accessibility - the ability of an authorized user to access the required information service at allowed time.

In these models, information security issues required to be taken into account when setting up digital libraries are limited or do not exist completely. As mentioned above, adding security modules to already installed system is insufficient. Because some security requirements may require modifications to other modules and cause additional problems in the system integration. Therefore, it is necessary to look for formal models for secure digital libraries. This article offers the *IDEA4SP* model for structuring information security issues in digital libraries. The *IDEA4SP* (Infrastructure, Digital Environment, Authentication, Access Control, Authorization, Audit for Service and Privacy) model consists of the following components:

Infrastructure - covers communication and hardware security, network security and web security issues in digital library environment.

Digital Environment - implies security of the digital content of the digital library that is the data and metadata.

The following *4A* security services include:

Authentication - ensures the second party that a person is verified. The user may prove his/her authenticity by submitting at least one of the following sources: what he/she knows (password, personal identification number, cryptographic key), what he/she owns (personal card or other similar device) or whatever he/she has (biometric characteristics as voice, fingerprints).

Authorization – enables the user access by defining the transactions (access rights) to be performed by the users over the digital objects.

Access control - prevents unauthorized use of the resource (i.e. controls who has access rights, under what conditions the access is allowed, and what the user accessing the resource is allowed to do). Different access control models are available and each of them can be used [20].

Audit – performs recording, collection and analysis of information about the events taken place in the digital library information system. It ensures the accountability of the users and the digital library staff and detects the attempts to violate the information security.

Service - describes how digital library actors behave and perform their digital library services and covers the security issues of digital library services.

Privacy – provides the confidentiality of the personal data of digital library users and their use with the purposes declared only to the user [21, 22].

Information security threats in digital libraries

Web security issues at European Library websites are explored in the study [23]. Using web security scanners, the websites of 80 libraries from four countries are analyzed in terms of security gaps. The analysis shows that most web sites have serious security gaps. The study found out that, although the legislation of the countries required the security of the sites, the libraries did not take appropriate measures to ensure the safety of online information systems. Several certain examples of the methods are demonstrated to improve the safety [23].

Session seizure is also one of the threats, where the user uses a session to get unauthorized access to information or service. This is mainly implemented by stealing the *HTTP "cookies"* (data files) used for the authentication with the remote server. Authentication here is an important mechanism; however at the beginning of the session it is not enough to authenticate the user. Authentication should be maintained throughout the session. The encryption of traffic between the communications objects is the mechanism to prevent the seizures (e.g. using the *SSL* protocol).

Digital content publishers and providers must have certain access rights assigned to each digital object. The underlying security threat is the disclosure of content, namely, the violation of the right to access the digital object. The confidentiality of this data should be ensured [20]. Data encryption is a mechanism used to provide confidentiality.

Another security threat is data modification. Access rights may not allow certain users to modify the content of a digital object. In this case, any modification of the content will be the violation of the designated access rights. Therefore, in order to detect the modification of the digital object, the data integrity controls must be guaranteed to be assured that they are imported without any modification by the authenticated entity [20]. There are various methods to ensure the integrity of the data, some of which are hash functions and data authentication codes (*e.g. message authentication codes MACs*).

Unauthorized disclosure of the digital content and its metadata in the digital library is also a security threat. To avoid this attack, each of the above services can be independently or jointly used. Authorization is the preferred service for the digital library [24, 25].

Another security gap is the catalog data modification. Catalog data contains all the information about the metadata; any change in its content may cause numerous problems. The attacker may change the access rights of certain documents and cause infringement of the access rights specified by the resource owner. Therefore, the data integrity must be monitored to detect changes to the catalog data. This can be implemented by using hash-functions or *MAC* codes.

Unauthorized data use is another attack. An attacker can get access by legitimate ways and then exploit his/her privileges by misusing the data, for example by copying the document. This attack can be prevented by controlling access and protecting the intellectual property rights. The term "*digital rights management" (DRM)* is used to describe the intellectual property rights and the protection of content from various logic security attacks [26]. To provide *DRM*, a technology group, such as encryption, passwords, digital signatures, and watermarks is used [27].

Avoiding authentication is also a threat. The attacker can perform certain actions intended for only authenticated users. This may result in disclosure of protected information or modification of data. *Kerberos* and *X.509* authentication services can be used to prevent avoiding the authentication. Thus, they test the users' identifiers before the data exchange starts [20].

DoS attacks at the service level are one of the main threats. They aim at preventing the delivery of digital library services to the users. Accessibility is one of the three main aspects of information security. It is a system property that allows the authorized users to access the system and use its services anywhere and anytime. There are different types of *DoS*-attacks. As the reversing mechanism, the network administrator identifies the *IP* address of the attacker and disables his/her access to the network.

Masquerade is one of the authentication attacks, when the attacker falsifies his/her identity and presents himself as someone else. This is a key problem, because some users may have access to some content and others may not. If the attacker is introduced as a user with access to certain content, then the access rights to the object is violated. Authentication checks the authenticity of the subject involved in the communication and thus prevents the masquerade. *IP spoofing* is a type of low level masquerade (at network-level of *TCP/IP* model). In the course of *IP spoofing* the network packages with fake *IP* address are generated, and it is attempted to gain access to the confidential content. Authentication via packet filtering can prevent *IP spoofing*.

Another attack includes the misuse of the privileges. In this case, the user infringes the access rights to the digital object. This can be prevented by the access control mechanisms [28].

One of the problems in e-commerce environment is the denial of the communication fact by the user [29]. Some digital libraries may require payment for access to certain information. The payment process should be safe and prevent any avoiding behavior by the users. Non-avoiding is one of the welcomed requirements, which can be provided with digital signatures of the user for each transaction. The communications during a payment may be processed by an authorized third party (such as *PayPal).*

Security of personal data in digital libraries

Users willingly submit their names, email addresses, telephone numbers and other identifiable personal information to access digital library services. This information should be protected from the hackers and others who want to use this information for non-library purposes. Except the demographic data of the customers, the digital library has the access to the following personal information [30]:

- membership files;
- temporary e-books and information about the e-publications;
- online search information;
- email logs and other Internet activity;
- information about visited and downloaded web pages;
- user profiles for broadcast services;
- user navigation profiles;
- information requests.

Thus, digital libraries can collect and archive a large amount of information about their users. This database is often confidential between the library and the individual. However, commercial organizations, law enforcement agencies, power structures and libraries are interested in this information for services marketing. The studies show that when using libraries, the users' concerns about the confidentiality are often low. They believe that the library will not deliver their personal information to other organizations. Though the librarians principally respect the confidentiality of personal data as their professional value, they do not pay attention to this enough. Moreover, most libraries are not ready for data protection. Based on the proposals of the respondents, the main principles of the confidentiality policy have been identified [30].

The American Library Association specified their requirements to the personal data security in the Code of Ethics in early 60s [31]. Article 3 of the Code of Ethics sets out broad responsibilities (edited in 1995): "We protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted."

The American Library Association published "Questions and Answers on Privacy and Confidentiality" in 2002 and updated it in 2014 [32]. The document highlights the expectations and best practices in how libraries and library systems can process the confidential customer data. This document is written at high level, covering all types of library functions (e.g. staff rules, privacy audits, confidentiality on shared computers). However, the confidentiality of personal data in *OPAC* or integrated library systems is almost unclear.

Awareness of the librarians about the personal data security may have a positive impact on the prevention of personal data security violations. This assumption is assessed by the surveys before and after the training of the librarians on personal data [33].

Digital Rights Management

One of the problems caused by digital libraries is the insufficient copyright protection. *DRM* provides property right protection of the content through the content encrypting and associating it with the digital license. The license identifies the user authorized to access the content and displays the content list of the product and expresses the user's rights to the resource in the machine readable format. *DREL* and *XrML* languages are used to indicate these rights, restrictions and conditions [27].

The following technologies and mechanisms are used to ensure the protection of digital rights:

Encryption. Symmetric and asymmetric encryption methods can be used for access control. Open-key encryption is used in payment systems, which controls who uses the content and how.

Passwords. The sequence of characters saved in the user log system should coincide with the sequence of characters provided by the logging user.

Watermarks. Characters or photos are added to show the property right. Steganographic methods are used to place these data within audio, video or photos.

Digital signature. Typically, digital signature is realized with open-key cryptography in practice. Hash functions are used in the process of creating a digital signature.

Digital fingerprints. Digital fingerprint is a stronger method using digital signature and watermarks together. The content creator creates a unique, personalized, customized copy for each user. Therefore, they are called "fingerprints". The content creator can use search engines to find these copies.

Backup discovery systems - search engines can be used to detect such backed items. Copy detecting browsers can also assist the protection of the digital content.

Payment systems - some kind of security technology as it requires user registration, credit card authentication, and obliges the trust between the content provider and the customer. Installation of payment systems can assist the protection of the digital content.

There is no standard mechanism for *DRM* provision. One of the main reasons for this is the lack of regulation. However, there are various systems and protocols to provide the content management and support the fair use policies in this area. Thus, digital rights management is one of the most important and controversial issues where discussions between researchers and practitioners are far from consensus.

Using *RFID* technology in digital libraries

RFID (Radio Frequency Identification) is also widely used in libraries. *RFID* automatically identifies the objects, defines their locations and monitors their movements [35].

Each publishing unit (books, journals, etc.) in the library is embedded with a special *RFID* label that receives and transmits radio signals. They receive a unique electronic identification number with this label. The special device (reader) uses this label when receiving, transmitting, sorting, inventoring, and locating the publishing unit.

With *RFID* technology, the libraries can identify the location of the publishing units, facilitate the issuance and acceptance of books, accelerate the inventory process, prevent theft and change of literature, and perform delivery and receipt of the books without the participation of a librarian.

The use of *RFID* in libraries creates a number of privacy concerns [36, 37]. In addition, the information collected on the *RFID* can be exploited by malware through listening, tracking, forging, repetition, and service rejection attacks [38]. *RFID* technology attracts the attackers due to the initial software code, common protocols and multiple devices, databases, valuable data, and incorrect security considerations.

Three types of malicious software identified by *RFIDs* are *RFID*-exploits, *RFID*-worms and *RFID*-viruses [39]. The malware writes exploits to the *RFID* tags, and then the exploits infect the background software of the reader when it is read. This attack typically is targeted at low-cost *RFID* tags and non-contact smart cards.

RFID worms spread themselves on the network, and target *RFID* tags, emails, and files. They exploit security gaps in *RFID* services online.

RFID viruses require infected tags to spread the virus. The databases are attacked by these viruses.

Conclusion

The digital library is an open-access online information system, providing an unlimited number of readers with quality and comprehensive information anytime and anywhere. Distributed information resources of this type of library, which are locally and remotely accessible to the public, consist of both the digital library units owned by the library, and electronic editions acquired by the library either on a paid or exchange basis. The use of these electronic documents

of various types, category and contents with the safe, secure and intellectual property rights over the Internet is of great importance. Moreover, confidentiality of the personal data of the library users should also be protected against a variety of threats. To this end, the article analyzed the current state of scientific and practical research on information security and confidentiality in digital libraries and identified a numerous topical research areas.

References

1. Arms W.Y. Digital Libraries. – Cambridge, MA: The MIT Press, 2000, 304 p.
2. Al-Suqri M. Afzal W., Digital age: Challenges for libraries // Information, Society and Justice, 2007, vol. 1, no. 1, pp. 43-48.
3. Fox E., ElSherbiny N. Security and digital libraries. Digital Libraries - Methods and Applications. K.H.Huang (Ed.), InTech, 2011, pp. 151-160, http://www.intechopen.com/books/digital-libraries-methods-and-applications/security-and-digital-libraries
4. Anday A., Francese E., et al. Information security issues in a digital library environment: A literature review // Bilgi Dünyası, 2012, vol. 13, no. 1, pp. 117-137.
5. Hadow K. Data security for libraries: Prevent problems, don't detect them // Feliciter, 2009, vol. 55, no. 2, pp. 50.
6. "State Programme on Development of Library-information spheres 2008-2013" in the Republic of Azerbaijan 6 October, 2008, http://www.e-qanun.az/alpidata/ framework/ data/15/c_f_15493.htm
7. Khalafov A.A. Development of library study in Azerbaijan in the years of independence // Scientific works of ANAS CSL, 2010, No11, pp.3-31.
8. Aliyeva-Kangarli A. Development of electronic libraries: international and Azerbaijani experience, problems, perspectives // Scientific works of ANAS CSL, 2004, No 5, pp.3-11.
9. Jafarov C.A. Electronic catalog in library information service. Baku: Progress, 2012, p. 224
10. Alguliyev R.M., Mammadov E.Ch. Some Issues of Interaction of Integrated Library Systems and Electronic Libraries // ICT in Education, 2011, No3, pp.4-10.
11. Karimova S.H. Formation of commercial text databases, features and problems // Problems of information society, 2012, No2 (6), pp. 64-74.
12. Lancaster F.W. Libraries and librarians in the age of electronics. Arlington, VA: Information resources press, 1982, 229 p.
13. Antopolsky A.B., Vigursky K.V. The Concept of Electronic Libraries // Electronic Libraries, 1999. Vol. 2. Issue. 2, http://www.elbib.ru/index.phtml?page= elbib/rus/journal/1999/part2/antopol
14. Kruk S., McDaniel B. Semantic Digital Libraries. Springer Verlag. 2008, 245 p.
15. Baruzzo A., Casoto P., Challapalli P., Dattolo A., Pudota N., Tasso C. Toward semantic digital libraries: Exploiting web 2.0 and semantic services in cultural heritage // Journal of Digital Information, vol. 10, no. 6, 2009, http://www.journals.tdl.org/jodi/index.php/jodi/article/view/688/576
16. Lagoze C., Davis J.R. Dienst: An architecture for distributed document libraries // Communications of the ACM, 1995, vol. 38, no. 4, pp. 1.
17. Candela L., Castelli D., Ferro N., Koutrika G., et al. The DELOS Digital Library Reference model. Foundations for digital Libraries (Version 0.98), 2008, http://www.eprints.port.ac.uk/4104
18. Gonçalves M.A., Fox E.A. 5SL – A language for declarative specification and generation of digital libraries / Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, 2002, pp. 263-272.
19. Gonçalves, M.A., et al. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries // ACM Transactions on Information Systems, 2004, vol. 22, no. 2, pp. 270-312.

20. Online Computer Library Center (OCLC): OCLC Digital Archive Preservation Policy and Supporting Documentation. Dublin, Ohio, 2006, 19 p.
21. Stallings W. Cryptography and Network Security. 4 ed. Pearson Prentice Hall: 2006.
22. Neuhaus P. Privacy and confidentiality in digital reference // Reference & User Services Quarterly, 2003, vol. 43, no. 1, pp. 26-36.
23. Saeednia S. How to maintain both privacy and authentication in digital libraries // International Journal on Digital Libraries, 2000, vol. 2, no. 4, pp. 251-258.
24. Kuzma J. European digital libraries: Web security vulnerabilities // Library Hi Tech, 2010, vol. 28, no. 3, pp. 402-413.
25. Adam N.R., Atluri V., Bertino E., Ferrari E. A content-based authorization model for digital libraries // IEEE Transactions on Knowledge and Data Engineering, 2002, vol. 14, no. 2, pp. 296-315.
26. Ferrari E., Adam N.R., Atluri V., Bertino E., Capuozzo U. An authorization system for digital libraries // The VLDB Journal, 2002, vol. 11, no. 1, pp. 58 – 67.
27. Tyrväinen P. Concepts and a design for fair use and privacy in DRM // D-Lib Magazine, 2005, vol. 11, no. 2, http://www.dlib.org/dlib/february05/tyrvainen/02tyrvainen.html
28. ElSherbiny N. Security in digital libraries. Masters Thesis. June 2011.
29. Tolone W., et al. Access control in collaborative systems // ACM Computing Surveys, 2005, vol. 37, no. 1, pp. 29-41.
30. Information Supplement - PCI DSS E-commerce Guidelines. January 2013. 40 p.
31. Sturges P., Davies E., Dearnley J., Iliffe U., Iliffe U., Oppenheim C., Hardy R. User privacy in the digital library environment: an investigation of policies and preparedness // Library Management, 2003, vol. 24, no. 1/2, pp.44-50.
32. American Library Association Code of Ethics, http://www.ala.org/advocacy/proethics/codeofethics/codeethics
33. Questions and Answers on Privacy and Confidentiality, http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/qa-privacy
34. Noh Y. Digital library user privacy: changing librarian viewpoints through education // Library Hi Tech, 2014, vol. 32, no: 2, pp.300-317.
35. Gibb F., Thornley C., Ferguson S., Weckert J. The application of RFIDs in libraries: an assessment of technological, management and professional issues // International Journal of Information Management, 2011, vol. 31, no. 3, pp. 244-251.
36. American Library Association. RFID in libraries: privacy and confidentiality guidelines, 2006, http://www.ala.org/advocacy/intfreedom/statementspols/otherpolicies/rfidguidelines
37. Kelly E.P., Ericson G.S. RFID tags: Commercial applications v. privacy rights // Industrial Management and Data Systems, 2005, vol. 105, no. 6, pp. 703-713.
38. Ngai E.W.T., Moon K.K.L., Riggins F.J., Yi C.Y. RFID research: An academic literature review (1995-2005) and future research directions // International Journal of Production Economics, 2008, vol. 112, pp. 510-520.
39. Rieback M.R., Simpson P.N.D., Crispo B., Tanenbaum A.S. RFID malware: Design principles and examples // Pervasive and Mobile Computing, 2006, vol. 2, pp. 405-426.