Data Content Protection for Virtual Libraries Heritage

Mihai DOINEA, Lorena BATAGAN
The Bucharest University of Economic Studies
mihai.doinea@ie.ase.ro, lorena.batagan@ie.ase.ro

This paper presents aspects of digital content protection in virtual library systems. The legislation aspects are presented to better emphasize the need of new security mechanisms. Integrated library systems architecture is presented with focus on their main purpose, manipulating and rendering digital content to end-users. The cultural heritage stored in such systems is an important asset that needs to be protected against malicious manipulation. The characteristics of a smart virtual library, supported by an integrated library system, are analyzed and a security model is proposed for implementation, based on its particularities. The model aims to be an interface between the interactions of anonymous users with the Online Public Access Catalog of a virtual library that assures the protection of digital content. Conclusions are drawn to better support the idea of cultural persistence by the use of Information Systems.

Keywords: *Security, Virtual Libraries, Smart Society, Digital Rights Management, Content Interaction*

1 Introduction

Integrated Library Systems, known also as ILS, have become ubiquitous in the digital and smart library sector because of their ability to manage digital content in almost each stage, from acquisition till archiving. With the use of an ILS, librarians are working efficiently and more rigorously, having at their disposal correction capabilities that can influence to the better the results stored in the system.

The Digital Rights Management, DRM, applied within Integrated Library Systems for protecting the cultural heritage of a virtual and smart library can be supported by complex existing instruments. This approach is meant to support the idea of cultural persistence by the use of Information Systems in a smart society. Virtual library patrimony must be preserved and not denatured by illegal manipulation of malicious users, for this reason security measures must be applied.

The paper presents current research in the field of integrated library systems and digital content security by use of digital rights management systems. In chapter 3, aspects concerning the legislation of digital rights management systems are debated and the necessity of such systems is strengthened. Chapter 4 presents the characteristics of a virtual smart library supported by an integrated library system and how electronic resources are managed in such systems. Chapter 5 describes how DRM techniques can be used in order to protect digital content from illegal manipulations of external users.

The final part of the paper presents the findings regarding how security techniques are used through DRM support to achieve digital content protection. Arguments are submitted in support of this approach and new directions for improving such systems are proposed.

2 Literature review

Digital Rights Management Technology is widely used for protecting digital content released by different publishers. As mentioned in [1] due to the fact that now many users have access to a fast internet connection and also because digitization is a process that nowadays is easily to implement, huge amounts of digital content can be easily accessed through different sorts of portals. This becomes a security issue, adding fuel to the fire of piracy. Legislation has kept pace with the new threats and the software development industry counteracted with specialized tools, based on Digital Rights Management techniques. From an Integrated Library System point of view the problems are more easily to depict but the solutions are not always so easy to implement. An ILS manages digital content, humanity

DOI: 10.12948/issn14531305/20.3.2016.05

cultural heritage, which must be protected not only from threats that can affect its integrity but also from a digital rights perspective.

Cultural heritage in digital form is also the product of the library, the institution who creates it. Besides the value given by the intellectual property of the author, a digitized material has also an added value given by its creators. Lots of library institutions have specialized equipment run by well qualified personnel, in digitization laboratories, even a separate department that takes physical papers, books especially old ones and digitize them, all this process taking many hours of work and innovation. These steps represents a real support for a smart society. Virtual and smart libraries can improve the access and the use of books, journals and other types of multimedia content. In a smart society the information continues to change rapidly and user expectations increase every day. The libraries content has become more numerous and more complicated. Virtual and smart libraries made available, in real time, more information. In a smart society the readers have greater expectation from libraries.

In order to enrich an electronic version of a book or journal, librarians must use multiple tools that allow them to scan, extract text, mark tables and figures, and enhance each part of the material with separate metadata in order to pursue its final objective, that being a digital version that can be easily accessed and read. As presented in [4] all the metadata are added to the digital format in order to enhance discoverability and interoperability of the material. The annotations come as a semantic layer on top of the digitized version enriching its value with the assiduous work of librarians. In the end the cultural patrimony will have another digital masterpiece to be added to the digital cultural heritage content to improve the end users' experience.

Due to the fact that most of the virtual libraries are meant for open access and not by means of a subscription of any sort, access to vast amounts of digital material is very easy. For this reason a data protection schema is good to be implemented in an integrated library system in order to keep track of the material which is downloaded from a virtual libraries being spread across the World Wide Web without any trace. A new security layer must be present in an integrated library system with the main objective to protect the digital material that libraries are struggling to create and distribute. A security layer based on digital rights management methods would prevent the use of digital materials without copyright infringements.

3 Legislation framework of DRM

The legislation aspects must clearly emphasize the border between piracy and correct utilization of digital content. From this description comes the need of implementing security mechanisms and one can determine what types of protection systems are appropriate and what security characteristics need to be preserved.

A digital library basic operations targets mainly the digital content which is cultural material found in electronic format. In an ILS system, digital content is only protected by the mechanisms found at the database level, which assures the integrity of its data. The primary purpose of the digital content is to be presented through a web interface to anonymous users for access or download. Due to this application, the digital content must also be protected against other forms of manipulation that are not desired.

This type of limitations must be explicitly mentioned by the legislation and specialized software must be developed in order to enforce the legislation rules accordingly. Whereas DRM techniques mainly deal with protecting digital material from illegal manipulation of any kind, another path in DRM has been tackled with interesting results. Besides the normal approach in which digital rights management must prevent people to see content that they don't have access to, and to prevent copyright infringement, a new and maybe efficient way is to change consumer expectations about what they are entitled to do with digital content as described in [9]. This is a different view on digital content from an end user perspective. Changing consumer expectations about what they are or not entitled to

DOI: 10.12948/issn14531305/20.3.2016.05

do with digital material, might trigger the appearance of different types of content that can't be manipulated so easily by infringers.

Not only that DRM systems protect the digital content but they also control the entire distribution chain of multimedia material. For this reason the commercial implications that resides in the ownership of the material and on the revenues generated by it, determine several types of DRM systems from which we mention two important types:

- DRM systems that prevent illegal manipulation of digital content;
- DRM systems that form complex business models in which digital content can be accessed on a pay-per-view basis.

Due to complex implications that DRM systems trigger detailed legislation provisions must be included in the law in order to protect both the consumers and the stakeholders. Many countries enacted in their own laws and regulations the aspects concerning digital rights management systems, digital authorship or ownership concepts.

From a technological perspective, a DRM system, as stated also in [10], must have the following two distinctive roles that must complement each other, as follows:

- access control – control access to digital content and allow only authorized users to make use of it;
- usage control – protect rights of the copyright owner by limiting what an authorized user can do with the digital content.

A moral approach to DRM systems is discussed in [8], reflecting the need of privacy for users that are using digital content distributed through DRM systems. This paper raises the problem that policymakers and DRM developers must answer the question according to which the privacy invasion caused by DRM restrictions is legally cognizable. They must insert privacy security controls into their DRM systems in order to be safe for well-intentioned users.

An important aspect regarding the content homogeneity in DRM system is debated in [3]. The problem of accepting digital content arrived through different communication channels in the same digital rights management system is very challenging. Yet this is another aspect that must be settled down by stakeholders in such a way that the consumers should be protected by the legislation.

Due to the fact that DRM systems are not yet 100% safety proof and that the law eludes lots of aspects that are still important for protecting digital content, usage contracts between the digital content providers and consumers have emerged that oblige consumers to use DRM systems under specific conditions as depicted in figure 1.

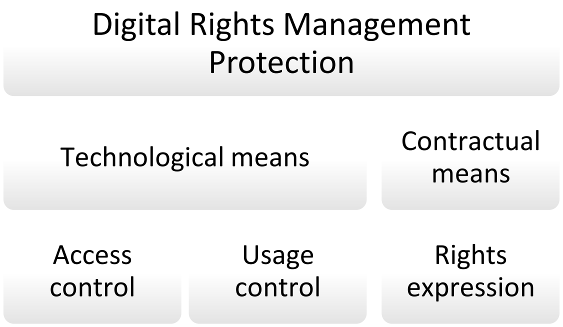

Fig. 1. Means of digital content protection through DRM

In [6] are presented ways of how digital rights management technologies are being used to limit consumer rights, also by use of contractual agreements to further narrow user rights. This makes the consumer to be part of a contractual agreement that will inflict penalties if users try to maliciously use the digital rights management system. This is a common practice for digital content providers, protecting in this way the digital content and also their DRM systems.

So DRM systems also make use, besides technological solutions, of contractual agreements for protecting their content. For this reason the digital content is bound by a metadata file, called rights expression languages, RELs that enables DRMs to read an entire list of usage rules. A very well-known implementation of RELs is the eXtensible rights Markup Language or XrML which describes the rights over a digital material such as: copy, delete, modify, execute, extract, embed, annotate, install, give, lease, print, display, transfer, loan, sell, play, restore, verify, save and many others.

DOI: 10.12948/issn14531305/20.3.2016.05

4 Virtual library systems

Integrated library systems together with other virtual library solutions are presented with focus on their main purpose, manipulating and rendering digital content to end-users. The cultural heritage stored in such systems is an important asset that needs to be protected against malicious manipulation. A list of information library system instruments are presented, developed by a specialized provider of library automation solutions, that can be interconnected in order to achieve information sharing for simplifying the end-users' access:

- Primo solution used to provide solutions for finding and obtaining a full spectrum of library materials, print, electronic, and digital, regardless of format and location;
- MetaLib used to allow searches on multiple databases;
- Rosetta is a tool used for preservation purposes due to the fact that a virtual library deals with large quantities of material;
- Aleph, an instrument for managing print collections, is an Integrated Library System that serves librarians in their effort to catalog the physical collections;
- DigiTool is a powerful solution used for displaying digital material in a more fashionable manner.

A classical ILS system implies the following actors and operations, depicted in figure 2.

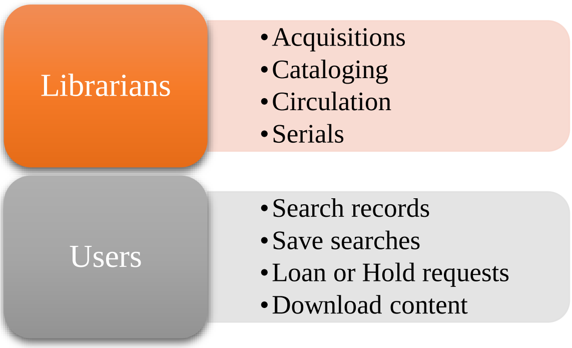

Fig. 2. ILS actors and operations

Integrated library systems are powered on a distributed system architecture which is very efficient in terms of resource consumption. But another generation of integrated library system emerged, offering even more advantages than the standard one. A new cloud based model, developed by Exlibris, brings a set of advantages that are hard to match by a distributed architecture. The instrument is an ILS called ALMA, successor of the current client-server product called Aleph.

ALMA supports library operations such as selection, acquisition, metadata management, fulfilment, digitization, resource sharing, search capabilities based on Apache Lucene engine and even the possibility of external integration into other types of library systems, according to [11].

ALMA has the advantage of being implemented in a cloud based architecture which offers great processing capabilities. If resources are no longer a problem than the entire attention is directed towards how digital content is manipulated and this is a very important aspect of any virtual library. ALMA not only processes digital content stored into the system to expose it to the public but it also does an opposite process, meaning that it is oriented towards user driven acquisition. This is a process based on which libraries will enrich their cultural patrimony by acquisitions based on titles that are requested by the users. Also digital content that is not frequently accessed is redirected to a specialized component used for preservation. In this way the system is kept clean and the search engine runs more efficiently. Another way of implementing this user driven feature, as presented in [11] is based on another important feature, the analytics engine, which, due to its multi-tenancy characteristic, can easily output relevant statistics about what materials are the most viewed and benchmarking about system performances.

An important aspect of an integrated library system that deals with huge amounts of data is to be as fast as it can when users send multiple search requests in the same time. Not only it must respond in a timely fashion but it also must give reliable results. For this the concept of key-queries is presented in [5], according to which a key-query is as dynamic content descriptors for documents and are defined implicitly by the index and the retrieval model of a reference search engine: key-queries for a document are the minimal queries that return the document in the top result ranks.

DOI: 10.12948/issn14531305/20.3.2016.05

The processes described in [5] assume that every document must have key-queries instead of the classical keywords. These key-queries will help search engines that are integrated in virtual libraries, to find much faster the documents that includes them. For determining the key-queries for a document in [5] is presented an exhaustive search algorithm along with effective pruning strategies, a decision tree machine learning technique.

One way of enriching a virtual library with interesting and appealing content is by digitizing the most requested titles on the market. Digitization involves library know-how and resource costs. For this reason, digitized material must be protected through special security controls. The digitization is a complex process that includes many stages, as depicted in the following picture, until the final product that will be displayed in the virtual library is obtained.

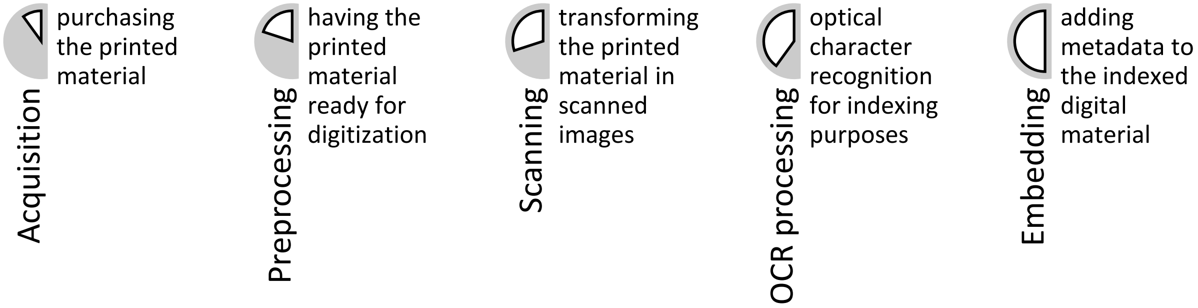
Acquisition
purchasing the printed material
Preprocessing
having the printed material ready for digitization
Scanning
transforming the printed material in scanned images
OCR processing
optical character recognition for indexing purposes
Embedding
adding metadata to the indexed digital material

Fig. 4. Digital content evolution in a virtual library

The digitization process starts from a printed material that is prepared for being transformed into an electronic resource. The preparation of the physical material consists in numbering each key element from the actual document like special descriptions, tables, figures and all types of elements that can't pass the optical character recognition phase.

After the document was prepared than the scanning process can begin, transforming the printed material into digital images that represent input for the following stage.

The process of optical character recognition is applied on each digital image and fragments of text are identified in order to be indexed.

The process of adding metadata for obtaining the final digital content uses special multimedia techniques that can embed data into the actual digital document or add it separately using special descriptors that are easily read by DRM systems.

For creating metadata in ALMA in order to connect digital content with associated records, the specialists make use of Digicorder using the Filemaker application. The application is used to describe the content of a book such as details of figures and tables, to number automatically the pages from the book, to add certain notes that appear in the original book and can't be easily reflected in the digital material.

The application also describes the chapters' structure creating an automatically content that serves as guidance for mapping the digital content reflected by the scanned images.

In order to facilitate content delivery to other library systems, a so called Metadata Interoperability Framework, MIF, can be used to share digital content across different platforms. The feeding process involves metadata definition, preview and validation feature, data push services based on Sword as well as OAI-PMH based data pull services and also mapping and transformation support.

5 DRM based data protection

The current implementations of virtual libraries systems are based on multiple software solutions integrated within a single system, allowing the manipulation of digital material at a central level. This type of systems focuses more on how to manage the material, how to make it searchable, how to display it in order to be more appealing to the end user.

DOI: 10.12948/issn14531305/20.3.2016.05

The protection of digital content lies only in the security mechanisms that are found at the storage level. For this reason a data protection configuration is analyzed for implementing it based on the particularities of a virtual library. The security instrument aims to be an interface between the interactions of anonymous users with the Online Public Access Catalog of the virtual library. A way of protecting the digital content from a library is to use specialized DRM techniques that restricts or limits the access to material based on access rights.

One way DRM acts for protecting the digital content is to transmit the information in an encrypted form to the consumer. The consumer devices is receiving the encrypted material and has the capability to decrypted on demand when the user tries to access using an authorized account to which it was granted access. This type of protection is called DRM through digital containers.

Encryption techniques have advanced because old ones were already hacked. New DRM model based on encryption techniques is proposed in [1] which will employ elliptic curve integrated encryption system (ECIES) and a secure one-way hash function for generating a dynamic one time content encryption/decryption key. According to [1], a portion of the key will be stored in license and the key will never be reused and stored on end user devices. This assures that successful attempts on getting the key and make it public will never end up in a successful attempt on accessing the digital content.

In [10] various means of DRM systems are presented. Their role is to protect or limit the access to the digital content by unauthorized consumers, such as:

- digital containers uses encryption for limiting access;
- rights lockers architectures used to allow users to access their own digital materials from different types of devices;
- copy generation management systems, or CGMS, limits the number of copies that a user can make from its digital content; after this limit the content becomes unusable.

DRM system face two major types of attacks that can allow a malicious user to access digital content without permission:

- attack for finding the encryption key; every user that has access to the encryption key is allowed to view and manipulate the digital content;
- unencrypted content capturing; for this attack users must exploit flaws in the standardization procedure of a DRM system.

In order to achieve protection of the digital material the DRM systems must use data for validation purposes, data attached to the original multimedia content. This information is known as metadata, data that describes a specific object, can be attached as a special header or even be embedded into the material by special techniques of digital watermarking or steganography.

Metadata enable reading of special sections of the digital content called descriptors that can contain information about:

- who provides the material in most cases called authorship;
- who is allowed to access in most cases called ownership;
- what types of operations are permitted based on the privileges assigned to the owner.

DRM not only can act as a protection layer for the digital content but it can also be used for commercial purposes in a pay-per-view architecture. In this way DRM systems are managing how the digital content is transmitted to the end users and how it is accessed.

There are advanced DRM systems developed that make use even of cloud based infrastructures, as presented in [7]. One of the reasons is that DRM systems must provide service to its users anytime, anywhere. For example UltraViolet ecosystem of DECE [12, 13, 14] supports the sharing of digital content between user devices as cloud-based digital authentication technique. In [15] was even proposed an architecture of DRM-as-a-Service that provides various functionalities of DRM as some services on the cloud environment, and we referred to it as the DRM Cloud.

DRM systems are combine also with trusted computing, [2]. This combination describes

DOI: 10.12948/issn14531305/20.3.2016.05

an open architecture for digital rights management enforcement on trusted computing platforms that empowers the consumer to select their operating-system and applications, including open-source options, without weakening the strength of the security functions. DRM techniques are used for a certain level of media objects that are sensitive and protected.

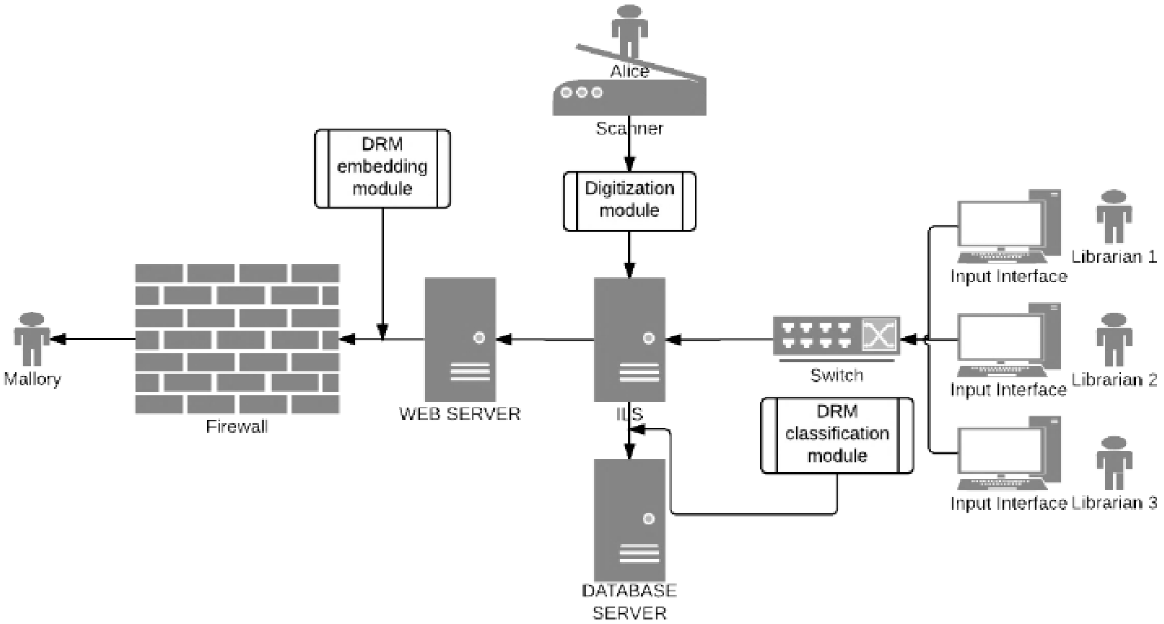

Fig. 5. Data model protection based on DRM in virtual libraries

The data model protection proposed in figure 5 is meant to be an interface between the interactions of Mallory, the malicious user with the digital content provided by Alice, the creator. The model aims to capture Mallory's requests for digital content and give him protected digital content according to a set of digital rights added for each object.

The embedding module of the presented data protection model runs automatically whenever a request is made to the web server for a digital object. The DRM classification module is triggered also automatically whenever new content is stored in the system.

6 Conclusions

Is very important that a virtual library system can expose its digital content, cultural heritage transposed in digital shape, with a certain level of security, due to its numerous interactions that are generating lots of vulnerabilities. Digital Rights Management mechanisms can be used to shape the interactions between end users and the digital content in ways that are meant to protect the digital cultural heritage of a virtual library system.

The DRM model is meant to add an additional layer of protection, restricting the access to original digital content without any kind of traceability.

Acknowledgements

Parts of this paper were presented at the 15th International Conference on Informatics in Economy, Cluj-Napoca, Romania, 2016

References

[1] K. Mehmood, M. Afzal, M. Mukaram Khan, M. M. Waseemiqbal, "A practical approach to impede key recovery and piracy in Digital Rights Management System," in *Proc. of 2015 12th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2015*, National Centre for Physics Islamabad; Pakistan; 13 – 17 January 2015, pp. 349-353

[2] A. Cooper, A. Martin, "Towards an open, trusted digital rights management platform," in *Proc. of the ACM Workshop On Digital Rights Management, DRM'06*. Co-located with the 13th ACM Conference on

DOI: 10.12948/issn14531305/20.3.2016.05

Computer and Communications Security, CCS'06, 2006, pp. 79-88

[3] W. Jonker, J.-P. Linnartz, "Digital rights management in consumer electronics products," *IEEE Signal Processing Magazine*, Vol. 21, No. 2, 2004, pp. 82-91

[4] M. Agosti, O. Conlan, N. Ferro, C. Hampson, G. Munnelly, C. Ponchia, G. Silvello, "Enriching digital cultural heritage collections via annotations: The CULTURA approach," in *Proc. of The 22nd Italian Symposium on Advanced Database Systems, SEBD 2014*, 16 – 18 June, pp. 319-326

[5] T. Gollub, M. Hagen, M. Michel, B. Stein, "From keywords to keyqueries: Content descriptors for the Web In SIGIR 2013," in *Proc. of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2013*, pp. 981-984

[6] A.S Kubesch, S. Wicker, "Digital rights management: The cost to consumers," in *Proc. of the IEEE*, Volume 103, Issue 5, 1 May 2015, pp. 726-733

[7] H. Lee, S. Park, C. Seo, S. U. Shin, "DRM cloud framework to support heterogeneous digital rights management systems," *Multimedia Tools and Applications*, 16 May 2015, 21p

[8] J. E. Cohen, "DRM and privacy," *Communications of the ACM*, Volume 46, Issue 4, April 2003, Pages 46-49

[9] P. Samuelson, "DRM {and, or, vs.} the law," *Communications of the ACM - Digital rights management*, Volume 46 Issue 4, April 2003, pp. 41-45

[10] S. Bechtold, "Digital Rights Management in the United States and Europe," *The American Journal of Comparative Law*, Vol. 52, No. 2 (Spring, 2004), pp. 323-382

[11] M. Doinea, P. Pocatilu, "Security of Heterogeneous Content in Cloud Based Library Information Systems Using an Ontology Based Approach," *Informatica Economica*, Vol. 18, no.4, 2014, pp. 101-110

[12] Kalker T, Samtani R, Wang X, "UltraViolet: Redefining the movie Industry?," *IEEE Multimedia*, 2012, pp. 7–11

[13] UltraViolet at http://www.uvvu.com

[14] Welcome to the UltraVilolet Wiki System Specification at http://www.uvvuwiki.com/images/3/3f/System-1.1r1.pdf , 2014

[15] H. Lee, C. Seo, S. U. Shin, "DRM Cloud Architecture and Service Scenario for Content Protection," *Journal of Internet Services and Information Security*, Vol. 3, No. 34, 2013, pp. 94-105

M. Doinea (1983) has been awarded with a PhD in the field of Economic Informatics, from The Bucharest University of Economic Studies (UES), Bucharest, Romania (2011). His PhD thesis tackles security optimization methods in the field of distributed applications. His entire research is focused on security topics, backed up by a master diploma in Computer Science Security (2008). Doinea started his professional career at UES, where he completed the entire educational track. Between 2013 and 2015, Doinea was affiliated with the Romania Academy from which he obtained a post-PhD scholarship in the field of Information Security. He is an associated professor at UES, teaching Data Structures, Advanced Programming Languages, Devices and Mobile Applications and Object Oriented Programming. He published many articles in collaboration or as a single author and his research interests are directed to areas such as computer security, mobile technologies, machine learning and vision algorithms.

DOI: 10.12948/issn14531305/20.3.2016.05

Lorena BĂTĂGAN has graduated the Faculty of Economic Cybernetics, Statistics and Informatics in 2002. She has become teaching assistant in 2002 and she has been lecturer since 2009. Currently, she is associate professor at the Faculty of Economic Cybernetics, Statistics and Informatics from the Bucharest University of Economic Studies. She holds a PhD degree in Economic Cybernetics and Statistics from 2007. She is the author and co-author of 4 books and over 50 articles in journals and proceedings of national and international conferences.

DOI: 10.12948/issn14531305/20.3.2016.05