Research on the Electronic Document Protection System with Responsibility Identification for Illegal diffusion

MIAO Quan-xing[1,a],Qin Yong-zhen[2,b*]

[1]Department of Information Engineering, Engineering University of CAPF, Xi'an Shaanxi, China

[2] Postgraduate Brigade, Engineering University of CAPF, Xi'an Shaanxi, China

[a]mpqx888@163.com,[b]qyz0513@sina.com

Keywords: Responsibility Identification; Digital Rights Management; Electronic Document Protection; Digital Fingerprint

Abstract. Current research situation in the electronic document protection based on Digital Rights Management technology is analyzed, the new model of electronic document protection system that can be available for identifying the responsibility for illegal Leakage of electronic document is put forward, and the key technologies in the model are discussed. By imbedding users' characteristics information regarded as digital fingerprint into the electronic document, the required technological means can be offered to identify the users' responsibility once the electronic document is leak illegally, so it has certain constraint for the users' illegal diffusion of the confidential electronic documents.

Introduction

A large amount of sensitive information in the network is in the form of electronic documents. Electronic documents are easy to copy and spread, so it is easy to cause the spread of confidential information which pose a threat to the security of confidential electronic documents. The main security threats exist in two aspects: one is unauthorized access and use of the user to the electronic document, the other is the illegal copying and dissemination of electronic documents by the legal users.

Rights Management Digital (DRM) is one of the hot topics in the field of information security, and it performs access control on digital information content in its survival cycle though a combination of hardware and software of the access mechanism. Its core is through a series of security technology to control the digital content and its distribution channels, so as to prevent the digital product to copy and use without authorization. At present, the research and application of DRM are mainly practiced in electronic books, Internet streaming media and electronic documents[1, 2]. DRM technology that is applied to the protection of electronic documents can effectively prevent unauthorized users from illegal access to electronic documents and use [3, 4].

About the problem of illegal dissemination of electronic documents by the legal users, the existing document protection system based on DRM is to solve the problem by restricting the user's copy of the original documents, such as the "save as" operation. But it is not suitable for the certain applications that legal users can to save the copy of the source documents and distribute the copy rightly. How to ensure the security of the documents, and at the same time the document can be effectively made use of by the legal users is the problem that this paper is to solve.

The basic idea of the model is that the digital fingerprint technology is combined with DRM in the system. By imbedding users' characteristics information regarded as digital fingerprint into the electronic document, the system may allow users to use the copy of the documents normally, at the same time identify the users' responsibility once the electronic document is leak illegally. As a result the system has certain constraint for the users' illegal diffusion of the classified electronic documents.

System Model

Considering the two aspects of the use of legitimate users and the protection of the document security, we combine the digital fingerprint technology with the DRM, and an electronic document protection model is designed, which may identify the users' responsibility once the electronic document is leak illegally.

The system is made up of five parts: the Document Distributing End, the Distribution Server, the Digital Fingerprint Server, the DRM Server and the Document Using End, as shown in figure 1. the Digital Fingerprint Server is made up of the Fingerprint Tracking System, the Fingerprint Encoding System and the Third Party. Different from the traditional electronic document protection system based on DRM, the Distribution Server and the Digital Fingerprint Server are increased.

The main work flow of the system is as following:

(1) The Document Distributing End generates a license of a document by. The license provides policy for the document management and protection, and after the identity authentication, the license will be submitted to the DRM Server.

(2) The Document Distributing End accesses to the Digital Fingerprint Server, to provide the user identity information to the Digital Fingerprint Server. The Fingerprint Encoding System in the Digital Fingerprint Server encodes the user identity information, and then sends the encoding sequence back to the Document Distributing End, the encoding sequence is embedded into the original document as the user's signature by the Document Distributing End.

(3) A copy of the document is transformed and encrypted at the Document Distributing End , and the copy is transformed to the DRM document and submit it to the Distribution Server. The establishment of the Distribution Server is to unify the release of the DRM document, eliminating the possibility of users from other ways to

obtain DRM documents. This is the need to achieve the responsibility identification of illegal distribution.

(4)The Document Using End\the user provides a valid user account and file name, if pass the identity authentication, the Document Using End access to the required DRM document.

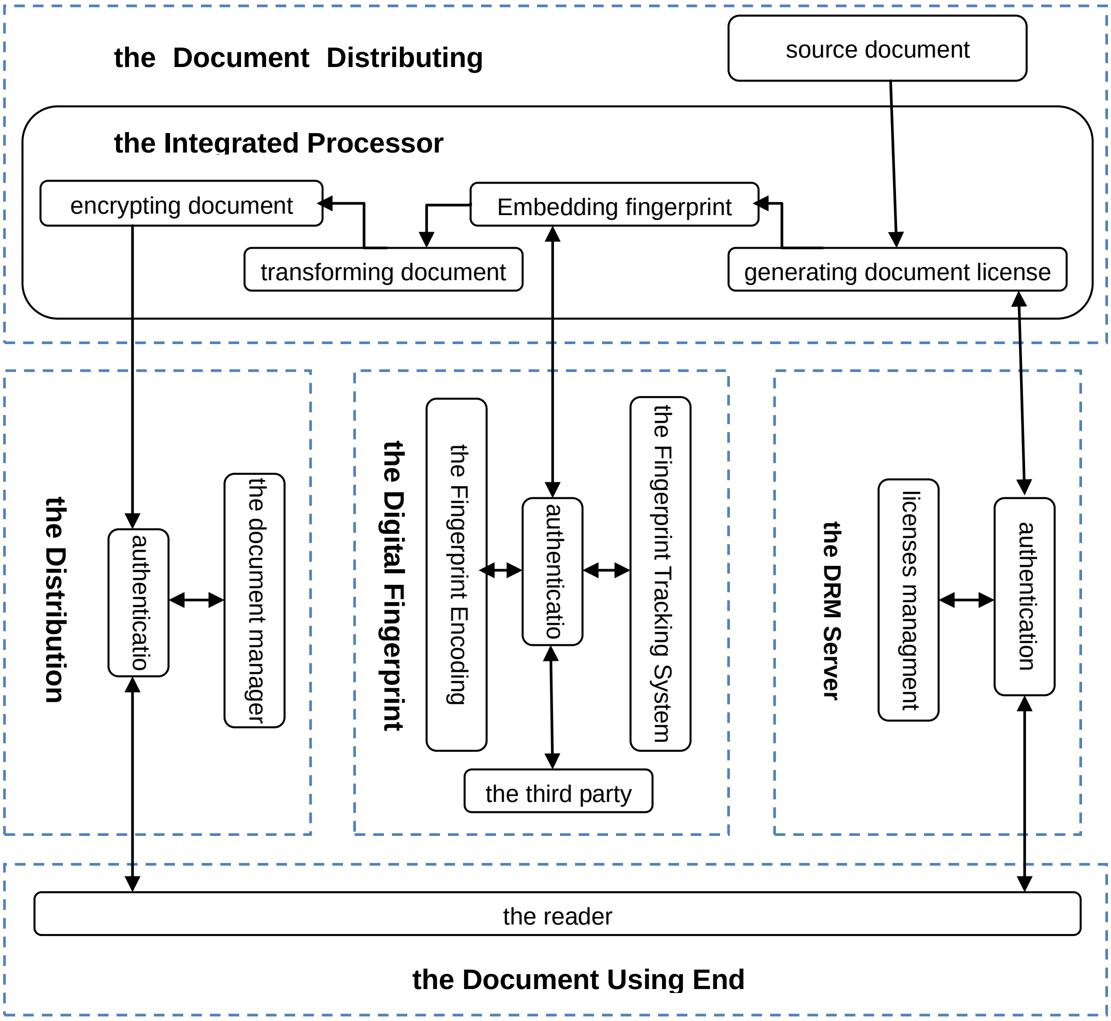

Fig. 1 System Model

License generation, document conversion, document encryption, DRM server and The Document Using End are same with the traditional electronic document protection system based on DRM. The main functions and implementation methods of the Distribution Server, fingerprint embedding, and the Digital Fingerprint Server are discussed in detail as following:

A. The Digital Fingerprint Server

The Digital Fingerprint Server is composed of the Fingerprint Encoding System and the Fingerprint Tracking System. In the meantime, The Digital Fingerprint Server completes the identity authentication with the third party together.

1. The Fingerprint Encoding System

The function of the Fingerprint Encoding System is fingerprint encoding, it is means that the Fingerprint Encoding System encodes the user ID to get the corresponding fingerprint sequence. Please refer to the following contents.

2. The Fingerprint Tracking System

The Fingerprint Tracking System mainly completes the extraction and matching of fingerprint. The system first extracts the fingerprint of the document, and then compares the extracted fingerprint with the fingerprint record in the database. Fingerprint extraction and matching process is: first using fingerprint extraction algorithm to extract the fingerprint sequence from the electronic document, decoding the fingerprint sequence for error correction to get the corrected fingerprint sequence, matching the corrected fingerprint sequence with the fingerprints in fingerprint database to make certain which user has copied and distributed the documents, and who should be responsible for. Tracking process is shown in figure 2.

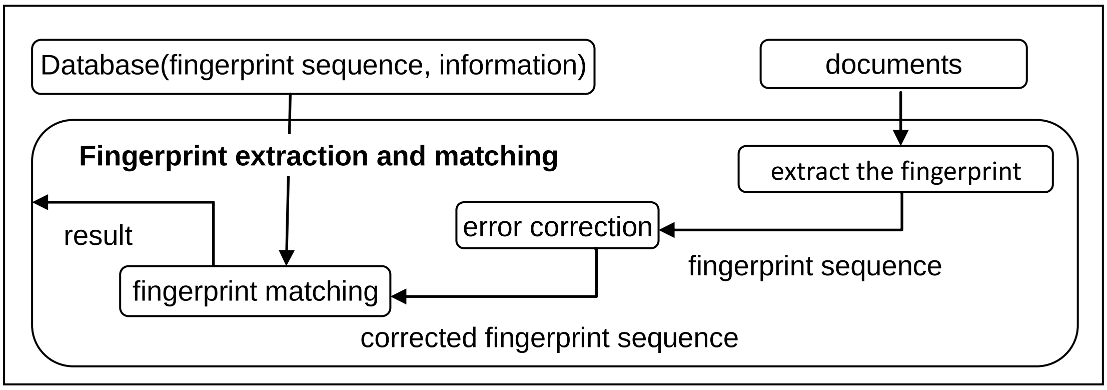
Database(fingerprint sequence, information)
documents
Fingerprint extraction and matching
extract the fingerprint
result
error correction
fingerprint sequence
fingerprint matching
corrected fingerprint sequence

Fig.2 Fingerprint Tracking Process

B. Embedding Fingerprint

The embedding fingerprint is to embed the fingerprint sequence of user code into a copy of the original document. As the basis for copyright tracking, the user code fingerprint sequence is embedded in each copy of the original document. Embedding process is: first, the information to be embedded is pretreated to get the user identity information fingerprint sequence that is the actual data generated from the user's name, work unit, public key and so on. Then the user identity information fingerprint sequence is cored into the formation of encoded fingerprint sequence. And finally, applying fingerprint embedding algorithm such as NEC algorithm, encoded fingerprint sequence is embedded into the document copy by the Document Distributing End.

The reason why the fingerprint embedding is placed in the Document

Distributing End is, if the fingerprint embedding is done by the Digital Fingerprint Server, the Digital Fingerprint Server may be too heavy to form a bottleneck of the whole system. In addition, transmitting a copy of the original document between the Document Distributing End and the Digital Fingerprint Server may also make some security problems.

C. Distribution Server

The distributor server is the core of the design and implementation of the system, which is mainly composed of user management subsystems and DRM document management subsystems.

1. User Management

The user management subsystem mainly provides the user management, and performs identity authentication when the user logs in. Administrators to establish a user database on the Distribution Server, which mainly store the user's account and password information. The user identity information on the Distribution Server can take the technical means to ensure the consistency of user identity information in the DRM Server and the Digital Fingerprint Server.

2. DRM Document Management

The DRM document is transmitted by the Integrated Processor to the Distribution Server through the secure channel, and the DRM document management subsystem saves and manages the document and the document information, such as the document encoding, document name, document user, document producer and date, etc.. Each document is relative to its legitimate users, so that the documents will be distributed to the appropriate user in the system.

3. Applying for the Documents

When a user needs to apply for a document, the following jobs should be done:

(1) The users log in the Distribution Server, and provide a document name or document code of the document after the identity authentication is successful;

(2) The Distribution Server queries in the document information database according to the document information entered by the user. If there is no the user identity information who can access the documents, the user requests are rejected; otherwise the user is allowed to use the documents, the Distribution Server takes out the document and send to the user which is marked with the user's information;

(3) Recording the information about document distribution such as the date of

issuance, the user identity information, etc.

Key Technologies

In the document protection system, the key technology is digital fingerprint technology, including the characteristics of the digital fingerprint algorithm and selection, the selection of fingerprint information and encoding etc.

As an important branch of information hiding, digital fingerprinting is mainly used for copyright protection. In digital fingerprint, a different symbol recognition code - the fingerprint is embedded into digital media, which is then distributed to the users. By embedding identity information of the legal users, the illegal copy of the digital media can be tracked, the responsible person for illegal media Distribution can be determined. By binding the user information with digital content, the digital fingerprint is one of the key technologies on information tracking and protection. So it is the effective means and one of the key technologies in the confidential electronic document protection [5, 6].

A. Fingerprint Algorithm

Most of the fingerprint algorithms are not very different in nature, and in the embedding process of those algorithms, original sequences are modulated with the masking property. The main fingerprint algorithm includes the spatial algorithm, the transform domain algorithm, the compressed domain algorithm, the NEC algorithm; the physiological model algorithm [7], each algorithm has its own characteristic. But because these algorithms are mainly aimed at image, video and audio data, it is not suitable for us to embed the fingerprint in the text. Up to now, several fingerprint algorithms for text have been proposed; they can be divided into two categories: one is based on the document image; the other is based on natural language processing technology. The fingerprint algorithm based on document image is divided into the document fingerprint algorithm based on the image block and the character characteristic fingerprint algorithm. The fingerprint algorithm Based on the image block can resist binarization attack, so it can resist the printing and scanning attack, but it is not considered the characteristics of a single character in the embedding process, so the capacity is low. The character characteristic fingerprint algorithm achieve the fingerprint embedding though modifying the character features of some of the characters in the document, this kind of method can adapt to the Chinese square character, through the proper choice of the embedded strokes and the modified character, to improve the capacity of fingerprint algorithm. This algorithm not only improves the flexibility of the system, but also increases the ability of the system to withstand the attack. It provides a better security and privacy for the information.

Therefore, it should be selected in the system to carry out the fingerprint embedding.

The purpose of the fingerprint tracking algorithm is to recognize the original user of the data copy after the data copy is illegally distributed. The copy of the data processed by the tracking algorithm may be processed by a single user attack or collusion attack, so a good tracking algorithm must have a certain ability to resist attack. In the case of a single user attack, users are likely to not be honest directly to the document illegal, and smart leak will be in the illegal dissemination of information, the use of various methods to deal with his data copy, to achieve the purpose of eliminating the fingerprint. It's more difficult to track the illegal spread of the way that it might make a copy of the way they have been created. Therefore, in the research of tracking algorithm, it is necessary to study the method of all possible attack, which can provide a reliable basis for the illegal behavior of the pirate [7,8]. Due to space constraints, the problem is not discussed in depth.

B. Fingerprint Encoding

To track the documents illegal copying done by legal users, digital fingerprint is introduced into document protection. Digital fingerprint embedded into the document must be difficult to be tampered or forged. Therefore, we choose ID as the fingerprint information and embed it in the electronic documents. User ID is generated by the user's name, work unit, user's public key together. A user's public key is a key that is used to encrypt a license when a DRM server provides a user's required license. The Digital Fingerprint Server, the Distribution Server and the user identity information on the DRM Server should be consistent, and the user ID should have the uniqueness in the entire management domain.

After selecting the fingerprint information, the user's fingerprint should be encoded to increase the ability of the collusion tolerance of the fingerprint scheme when the fingerprint information is issued [9]. Fingerprint information Encoding scheme is the process that how the user related information is encoded according with certain rules to generate the code words with a certain ability to resist the attack. A good fingerprint encoding is the key factor to track the illegal distribution, each fingerprint encoding scheme has a corresponding tracking system. At present, the encoding method based on text includes the following several: shift encoding method, synonym replacement method, feature encoding method, transform encoding method [10,11]. Each encoding way is different in the encoding.

Conclusions

In the paper, digital fingerprint technology is introduced into the document protection based on DRM electronic document protection technology, in the model a new way is proposed for the responsibility identification of confidential electronic document illegal copying and distributing and the key technologies of the model are studied. In order to realize the scheme, it also needs to do a lot of research on the robustness, security, and so on.

References

[1] Gelareh Taban,,Alvaro A. Cardenas, Virgil D. Gligor. Towards a secure and interoperable DRM architecture[J].Proceedings of the ACM workshop on Digital rights management,2006.

[2]Sai Ho Kwok. Digital Rights Management for the Online Music Business[J]. ACM SIGecom Exchanges,2002,3(3):17～24.

[3] Palgrave Macmillan.The use of a digital rights management system for article delivery [J].Journal of Digital Asset Management, 2006,2 (2) .

[4]Andrew Braid.The use of a digital rights management system in a document supply service[J]. Interlending&Document Supply,2004,32 (3):189～191.

[5]Kim M,Kim J,Kim K.Anonymous Fingerprinting as Secure as the Bilinear Diffie-Hellman Assumption.In: Proc.of Information and Communications Security: 4th International Conference,vol.2513 of LNCS. Berlin/Heidelberg:Springer-Verlag,2002.99～110

[6]Chung C, Choi S, C hoi Y, et al. Efficient Anonymous Fingerprinting of Electronic Information with Improved Automatic Identification of Redistributors. In: Proceedings of the Third International Conference on Information Security and Cryptology, vol.2015 of LNCS. Berlin/Heidelberg: Springer-Verlag,2001.210～239

[7]Birgit P, Ahmad-reza S. Anonymous Fingerprinting with Direct Non-repudiation. In: Advances in Cryptology-ASIACRYPT 2000: 6th International Conference on the Theory and Application of Cryptology and Information Security, vol. 1976 of LNCS. Berlin/Heidelberg,2000.400～429

[8]BRODER A Z. Some applications of Rabin's fingerprinting method[C]//Sequences II: Methods in communications, security, and computer science, New York: Springer-Verlag,1993:143～152.

[9]A Barg, G.R.Blakley,G.Kabatiansky. Digital Fingerprinting Codes: Problems Statements, Constructions, Identification of Taitors. IEEE Transactions on Information Theory,2003,49(4):852～865

[10]JADALLA A, ELNAGAR A. A Fingerprinting-Based Plagiarism Detection System for Arabic Text-Based Documents[J]. Intelligence and Security Informatics,2012,(7299):145～153.

[11]J.Lofvenberg,N.Wiberg.Random Codes for Digital Fingerprinting. Technique Report, LiTH-ISY-R-2059, Department of Electrical Engineering, Linkoping University,2000