Advanced Engineering Forum *Online: 2012-09-26*
ISSN: 2234-991X, Vols. 6-7, pp 452-458
doi:10.4028/www.scientific.net/AEF.6-7.452

A Buyer-Seller Digital Watermarking Protocol without Third Party Authorization

Xie Jianquan[a], Xie Qing[b], Tian Lijun[c]

Department of Information Management, Hunan University of Finance and Economics, Changsha 410205, China

[a]xiejianquan @sina.com, [b]csxieqing@126.com, [c]tian_lj@126.com

Keywords: Copyright Protection, Digital Fingerprinting, Asymmetric Fingerprinting, Fingerprinting Protocol, Piracy Tracing

Abstract. For protecting copyright of digital work, a kind of digital work commerce protocol is proposed based on RSA public key system. Digital fingerprint information that embedded in digital work can only be generated by the cooperation of the publisher and the user. The publisher cannot obtain this digital work, but the publisher can judge whether the fingerprint information exists in illegal copied digital work. This protocol can be utilized to trace and accuse the illegal piracy. Meanwhile, it can also protect legal user from framed by publishers. The digital fingerprint identification procedure that dispenses with participation of the third party can effectively avoid forgery in the procedure, which makes the scheme more convenient and safer than the scheme that needs the third party.

Introduction

The convenience of digital work duplication provides an easy way for pirates. As a result, digital watermarking and fingerprint technologies are widely used to protect multi-media digital works without decline of the quality of those works and awareness of the fingerprint. The digital watermarking is mainly used in identifying ownership, while the digital fingerprint is used in buyer identification. The latter method usually uses fingerprint as the only mark of buyer's identity. Publishers add this mark in the sold copy of digital work, by which they can trace the source of illegal copy when piratical duplication is found, and then pursue legal obligation of illegal duplication. In early period, digital fingerprint schemes only concerning with symmetric fingerprint plan, the emphasis of which is fingerprint coding research in preventing collusion attack of buyers. Researchers have developed many works on the field of withstanding the collusion attack [1-5].

Another problem that digital fingerprint facing with is that the symmetric scheme cannot determine the source of the illegal duplication after the piracy is detected. Does this piratical work is duplicated by a user, or it is duplicated by the publisher and then impute to other people? To solve this problem, scholars proposed a concept named asymmetric fingerprint and many kinds of buyer-seller protocol that base on the asymmetric fingerprint [6-8]. In this asymmetric fingerprint method, the digital work containing fingerprint requires users to firstly decipher the work with private key. The publisher has no way to obtain digital work containing fingerprint that can be directly used. By doing so, legal user can be effectively protected from malicious accusation of the publisher. At present, most of asymmetric schemes require the publisher offers evidence to an authoritative and credible third party when he accuses a user's illegal duplication. The third party, like certificate authority and copyright center [9-11], should apply valid authentication. However, if the third party colludes with the user, the user can get rid of the accusation, and can even obtain original digital content and then further infringe equity of the publisher. On the other side, if the third party colludes with the publisher, the publisher can frame up any legal user. The literature [12] proposed a kind of asymmetric fingerprint method without requiring the third party. But this method has low efficiency on discovering forgery in the fingerprint identification procedure, and it also requires the user to process many data. On the basis of the literature [12], this paper proposes a kind of buyer-seller protocol that does not need authoritative third party to access information. This protocol can

be used in piracy tracing and accusation, it can prevent that the publisher frames the user, and can also effectively avert user's forgery during the identification procedure. In addition, this protocol has significantly decreased calculation quantity compared with [12].

Asymmetric Fingerprint Scheme Based on RSA Public Key System

The RSA cryptographic algorithm is a widely used asymmetric cryptographic algorithm with highly security. The algorithm use keys to encrypt and decrypt, where the encryption key (k_d, n) is public, as it is called the public key, and the decryption key (k_e, n) is safeguarded by the user, which is correspondingly called private key. The encryption and decryption procedures are:

$$c = E_{k_d}(m) = m^{k_d} \bmod n \ . \tag{1}$$

$$m = D_{k_e}(c) = c^{k_e} \bmod n \, . \tag{2}$$

In the RSA encryption algorithm, n is the product of two big prime numbers. The security of the algorithm is mainly decided by the difficulty of disintegrating n. Hence it is normally required to guarantee that n is longer than 512 bits to ensure the security.

The asymmetric fingerprint scheme in [12] is a kind of buyer-seller protocol that based on RSA asymmetric encryption system. The sold products within this protocol contain digital fingerprints that used in tracing products and accusing illegal duplicator, respectively. The seller embeds the fingerprint in the production for product tracing before it is sold, and then the seller and the buyer generate the fingerprint using for illegal duplicator accusation together. The process to generate this copyright watermarking that embedded in digital product is as follows:

Step 1: When a user request to purchase a multi-media digital product, the publisher will firstly embed a tracing digital fingerprint DT in the original product to generate a product X for sale. Each product for sale has different tracing fingerprint. This fingerprint is only used to trace illegal duplication instead of serve as evidence in accusing the illegal duplicator.

Step 2: The user choose a digital fingerprint DF independently, here DF is a vector that its value is uniformly near 1. That is, every element of DF is uniformly valued in (1-ε,1+ε), ε is a trivial positive number. It means that $DF = \{df_i, i = 1,2,\cdots L\}$, $df_j \in \{1-\varepsilon, 1+\varepsilon\}$, where L is the sample amount (or coefficient amount in transformation field) of the digital work X. The user will keep the secrecy of DF, and use 256 bits to denote every element in DF, and then use the public key to encrypt elements individually:

$$DF^{'} = E_{k_d}(DF) \ . \tag{3}$$

After obtained encrypted chosen fingerprint $DF^{'}$, send this ciphered vector $DF^{'}$ to the publisher.

Step 3: After received $DF^{'}$, the publisher scrambles the sequence of elements in $DF^{'}$ and obtains DF^{*}.

$$DF^{*} = \psi(DF^{'}) \ . \tag{4}$$

The scrambling method ψ needs to be maintained secretly. The publisher use 256 bits to represent every element in X' as well, and encrypt every number of user's public key (k_d, n). Then multiply it with corresponding location of DF^{*} and then get vector X*, that is:

$$X^{*} = E_{k_d}(X') \times DF^{*} = E_{k_d}(X') \times \psi(E_{k_d}(DF)) = E_{k_d}(X' \times \psi(DF)) \ . \tag{5}$$

The publisher sends the vector X* as an encrypted product to the user.

Step 4: The user utilizes the private key (ke, n) to decrypt X* and then gains a digital product with tracing fingerprint. Because the length of X and DF are both 256 bits, the length of their product will not overpass 512 bits. Hence there is:

$$Y = D_{k_e}(X^*) = D_{k_e}(E_{k_d}(X' \times \psi(DF))) = X' \times \psi(DF) \quad . \tag{6}$$

From Eq.6, if only the value of each element in vector *DF* is close to 1, *Y* will close to *X*'. In other words, it will not affect the application of the multi-media digital work *X*', and it will not damage the tracing fingerprint *DT* which is embedded in the work neither. During this process, the publisher cannot get the production since the publisher does not know the private key of the user. On the other hand, the user can not erase the fingerprint *DF* as the user does not know the scrambling method ψ and consequently cannot obtain $\psi(DF)$. After the processes mentioned above, the user can get digital production with fingerprint embedded in. If the user only paid for use right rather than transfer right, the user is forbidden to copy this digital product. When a user who didn't pay for use right owns a protected digital product, the user that illegally duplicated the version he/she purchased can be confirmed according to the fingerprint *DT* embedded in the illegal duplication. However, *DT* cannot serve as evidence in accusation, and then an authorized third party is needed in that case. The publisher and the user provide the scrambling method ψ and *DF* to the third party respectively. The third party will decide whether *DF* is contained in the illegal duplication. If the third party provides the scrambling method ψ to the user, the user can erase the fingerprint. On the other hand, if the third party gives *DF* to the publisher, the publisher can generate *Y* and then frame up legal user. Consequently, both the publisher and the user hope to maintain the secrecy of their information in the fingerprint identification procedure.

Improved Asymmetric Fingerprint Scheme

Considering the defect of the mentioned scheme, a kind of asymmetric fingerprint scheme that does not need the authorized third party is proposed. The core idea is that collectively generate a batch of test signals by both the publisher and the user, only one of which contains the fingerprint *DF*. If the illegal duplication engendered by the user indeed, the publisher can obtain an approximate *DF* by compare it with the original content, and then the test signal that contains *DF* can be determined. No matter whether the judgment of the publisher is correct or not, the user does not need to reveal *DF*, while the publisher does not have to betray the scrambling method ψ. The entire process can be executed without the participation of the third party. If this procedure is applied in the court, the judicial officer, act as a bystander, can judge whether there is violation and record the final identification result.

According to the scheme proposed above, when the publisher discover an illegal duplication *Y*', as long as the duplication *Y*' is basically identical with *Y*, the publisher can determine from which user the duplication is generated through the pre-embedded fingerprint *DT*. Because of:

$$DF \approx \psi^{-1}(Y / X') \approx \psi^{-1}(Y' / X') \ . \tag{7}$$

The publisher can also obtain a fingerprint that approximate to the other fingerprint *DF*, which is generated by the user, from the illegal duplication. That is to say the publisher can indicate *DF* from multiple vectors, whose elements are uniformly valued from $(1-\varepsilon, 1+\varepsilon)$. In that situation, only if *DF* uniformly valued in $(1-\varepsilon, 1+\varepsilon)$, the publisher can accuse the user for the infringement without requiring user to offer *DF* and revealing scrambling method ψ.

The generative process of digital product with fingerprint is as follows:

Step 1: Similar with the Step 1 in the previous part, that is: the publisher embeds a tracing fingerprint *DF* in the original digital work and gain the version *X*'. This fingerprint does not serve as evidence for illegal duplicator accusing neither.

Step 2: The user generates $2 \times L$ secret vectors DF_2 that every element of them are uniformly valued in $(1-\varepsilon, 1+\varepsilon)$. Then cipher DF_2 in accordance with formula (3) and get DF_2', and then deliver DF_2' to the publisher. The publisher will randomly choose half of DF_2' and request the user to decipher it. If the deciphered data comply with uniform distribution in $(1-\varepsilon, 1+\varepsilon)$ indeed, the rest of DF_2' will be considered as real accusing fingerprint *DF*'. The user is demanded to sign the *DF*', otherwise the user is believed have violated rule. Hence the publisher could refuse to sell the corresponding multi-media digital work to the user.

Step 3: The publisher obtains *DF** by scrambling *DF*', and then uses 256 bits to represent each element in *X*', and then encrypts each number with the public key (k_d, n) of the user, at last, multiplies it with the corresponding location of DF^* and gains vector *X**.

Step 4: The publisher ciphers *X*' with user's public key and multiplies it with DF', which has scrambled in terms of formula (4). Then send the production that contains tracing fingerprint and accusing fingerprint to the user. The user can obtain a digital production *X** by decrypt the production.

When illegal duplication appears, the publisher can determine the illegal duplicated user according to the pre-embedded fingerprint *DT*. Then the publisher can accuse the suspected infringing user and launch the fingerprint identification process. The steps of the process are as follow:

Step 1: The user generates (2*K*-1)×*L* random data S that are uniformly distributed in (1-ε,1+ε), that is: $S = \{s_i, i = 1,2,\cdots(2K-1)\times L\}$, $s_j \in \{1-\varepsilon, 1+\varepsilon\}$, where *L* has the same meaning as previous, *K* is a integer denoting the entire number of fingerprints, only one of which is the real fingerprint. The bigger the *K* is, the less the possible for the publisher to frame up the user is, but the calculation quantity is bigger as well. Then the user individually represents every element of *S* by 256 bits and encrypts it with the public key, and then obtain *S*'. That is:

$$s_i' = E_{k_d}(s_i), i = 1,2,\cdots(2K-1)\times L \ . \tag{8}$$

After that, calculates Hash value *SH* of each element in *S*':

$$SH = \{sh_i, i = 1,2,\cdots(2\times K-1)\times L\} . \tag{9}$$

Where $sh_i = hash(E_{k_d}(s_i))$, $i = 1,2,\cdots(2K-1)\times L$. Then transmit *SH* to the publisher.

Step 2: The publisher receives and records *SH*, and then randomly pick *L* elements from *SH* to construct vector SH_K, while the rest elements are randomly divided into 2*K*-1 vectors that marked by SH_j, $i = 1,2,3,\cdots 2\times(K-1)\times L$. The publisher will inform the grouping situation to the user.

Step 3: According to the grouping situation of the publisher, the user generates (2*K*-1) random vectors from the random data yield in Step 1 with the same grouping method. By doing so, a random vector group *ST* is obtained:

$$ST = \{st_i, i = 1,2,\cdots 2\times K-1), \tag{10}$$

where st_i are vectors with *L* dimensions, and every element of it is obtained from *S*, while no overlap exists.

Step 4: The user utilizes *ST* and *DF* to generate (*K*-1) test signals containing fake fingerprint and a test signal containing real fingerprint *DF*.

$$t_i = st_{2i-1} \oplus st_{2i} \quad , \quad i = 1,2,\cdots K-1 \ . \tag{11}$$

$$t_K = st_{2k-1} \oplus DF \ . \tag{12}$$

Then scramble the order of those *K* test signals (not scrambling the inner elements in those signals). After that, convey those signals to the publisher and require the publisher to indicate which signal contains the real fingerprint *DF*.

If the publisher has obtained *DF* from the illegal duplication, the number of interrelations between *DF* and every test signals can be worked out. Since every element of *DF* are uniformly distributed in (1-ε,1+ε), if the test signal does not contain the real fingerprint, the interrelated coefficient will close to 0. On the other side, if the test signal does contain the real fingerprint *DF*, the interrelated coefficient will bigger than 0. Hence the publisher can indicate that the test signal corresponding to the maximum interrelated coefficient contains fingerprint *DF*. The publisher cannot estimate *DF*, in case that the publisher frames up the user. The probability of pointing out the correct test signal is only 1/*K*, and it is easy to see that when *K* is big enough, this probability will approximate to 0.

Step 5: The user discloses the former (2*K*-2) random vectors of the (2*K*-1) random vectors that generated in Step 3, and the encryption value st'_{2K-1} of the (2*K*-1)-th random vector st_{2k-1} as well.

Step 6: The publisher verifies whether the following three conditions are true:

(1) Whether the disclosed (2*K*-2) random vectors are uniformly distributed in （1-ε,1+ε).

(2) Whether st_{2k-1} belongs to *S* (by judging whether the HASH value of each element of vector st'_{2K-1} belongs to *SH*).

(3) Whether $E_{k_d}(t_K)$ is identical with $st'_{2k-1} \oplus DF'$.

If any of those three conditions is false, the user is believed infringing in the fingerprint identification. If the user indeed duplicated the digital production and the user didn't infringe in the identification process, then the user cannot disavowal the correctness of the fingerprint identification that pointed out by the publisher. Meanwhile, because st_{2k-1} is steal private, the publisher cannot calculate *DF* from t_k, which guarantees the security of *DF* and eliminates the possibility of the publisher framing up the user. Otherwise, if the publisher frames a user and executes fingerprint identification, even if the attempt fails, the publisher can obtain test signal *DF* that contains real fingerprint. This will much benefit the publisher for later defamation.

From procedures above, it is clear that if the illegal digital production is certainly duplicated by the user, the publisher can indicate the test signal that contains the real fingerprint; if the publisher calumniates legal user, then the success probability is only 1/*K*. In the fingerprint identification process, the publisher and the user respectively maintain the scrambling method and the fingerprint *DF* secret.

Performance Analysis

In the scheme proposed in this paper and [12], there exists one important premise, that is: both the fingerprint and the random data offered by the user that adopted in fingerprint identification must uniformly distributed in (1-ε,1+ε). If the random data that provided by the user in identification process severely diverges from (1-ε,1+ε), then the identification result will be severely affected. This will result the user escaping from the accusation. However, in the identification process, one cannot command the user announces all of the random data, because the publisher can deduce the fingerprint of the user and this may result the possibility that the publisher framing up the user. To solve this problem, in the identification process of literature [12], the publisher requires the user announces part of the data to verify whether the user's generated data satisfy the uniform distribution requirement. This requirement can investigate that whether the user tricked in the identification. Nevertheless, if the user due illegally duplicated the digital production, the user may yield some data that severely diverge from (1-ε,1+ε) to counterfeit data that truly uniformly distributed in (1-ε,1+ε). As mentioned formerly, one cannot demand the user to announce all of the random data, the possibility that the user successfully fakes still exists. Here assume that m is the number of random data that generated by the user, r is the number of random data that the user is required to announce, and then the probability for the user to successfully fakes in the identification process is:

$$P = C_{m-1}^{r} / C_{m}^{r} = (m-r)/m \ . \tag{13}$$

On purpose of prevent the cheat, the user should reveal those random data as much as possible. On the other side, to avoid the publisher frames the user, m must achieve a certain magnitude, which cause *r* becomes huge.

The probability for the publisher successfully denigrating the user is:

$$P_1 = [\Phi(\frac{\varepsilon^2\sqrt{L}}{1.17})]^K \ . \tag{14}$$

Where $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$ .When the digital work is a digital image of 512×512, on purpose of satisfying information hiding imperceptibility, ε must smaller than 0.055, while K must no smaller than 100. It can be seen that the amount of data that require the user to generate and process is quite large. Therefore only part of data is requested to reveal in [12]. Assume that 1/3 of the whole data is commanded to announce, the rest 2/3 of data is used for identification, according to Eq. 13, the probability for a user to successfully counterfeit in the identification is 0.667. So the literature cannot avoid this counterfeit effectively.

In this paper, the probability for a user to successfully counterfeit in the identification is:

$$P_2=\frac{(2K-1)\times L-(2K-2)\times L}{(2K-1)\times L}\approx 0.00502\ . \tag{15}$$

It shows that $P_2 << P_1$, the capability of preventing counterfeit of the algorithm proposed in this paper is better than that of [12]. This algorithm can effectively find user's counterfeit activity in the identification process, and all of random data that generated by the user can be used in the identification. Moreover, the literature [12] needs an additional part of data to judge whether the user has counterfeited. If the user reveals 1/3 data, then the random data that needs user to process is 1/3 more than that in this algorithm. Hence this algorithm can significantly reduce the amount of work for users.

Conclusions

A kind of asymmetric fingerprint scheme that free from the authorized third party is proposed in this paper. The published digital production has two fingerprints embedded in, the fingerprint for tracing and the fingerprint for accusing piracy. When the publisher accuses a user for illegally duplicated a protected digital production, the publisher can indicate the exact user that executed the piracy, and can also verify the credibility of the identification to the judge without participation of any authorized third party. During the entire procedure, the publisher and the user can maintain their information secret. This will not cause potential threat, and can effectively avoid the user's counterfeit in the identification process. A new idea is provided in this paper for equitable electronic transaction and digital fingerprint. It has theoretical and practical sense in intellectual property rights protection and other fields.

Acknowledgments

This work was supported by Educational Science Subject of Hunan 12-th Five-Year Plan under Grant Nos. XJK011BXJ008

References

[1] A. Lu. A novel video watermarking scheme against manifold attacks. *Journal of Image and Graphics*, 14:11(2009), 2205~2211.

[2] C.L. Hou, C.C. Lu, S.C. Tsai, *et al*.An optimal data hiding scheme with tree-based parity check. *IEEE Transactions on Image Processing*, 20:3(2011), 880-886.

[3] S.H.Liu, L. Han, H.X. Yao. Video watermarking algorithm for resisting collusion attacks. *Journal on Communications,* 32:1 (2009), 14-19.

[4] L.Zhang, Q.L.Li, D.F.Hu. Scalable anti-collusion digital fingerprinting scheme. *Computer Engineering and Applications*, 48:1(2012), 128-131.

[5] X.Y. Wang, S. R. Douglas. Robust Correlation of Encrypted Attack Traffic through Stepping Stones by Flow Watermarking. *IEEE Transactions on Dependable and Secure Computing*, 8:3(2011), 434-449

[6] Y. Zhu, Y.T. Yang, Z.Y. Ye. An asymmetric spread spectrum fingerprinting with buyer transparency [J]. *ACTA ELECTRONICA SINICA*, 34:6(2006), 1041-1047.

[7] D.F.Hu, Q.L. Li. Bandwidth efficient asymmetric fingerprinting scheme. *International Journal of Communication Systems*, 25:2(2012), 84-91.

[8] C.C. Chang, H.C. Tsai, Y.P.Hsieh. An efficient and fair buyer-seller fingerprinting scheme for large scale networks. *Computers & Security* , 29:2(2010), 269-277.

[9] Z. J. Deng, X. J. Lai, D. K.He. Digital rights management model to resell right. *Computer Engineering*, 2009, 35(20):20-22.

[10] X.S. Chen, L.G. Liu, Z.D.Lu. Design and Analysis of Digital Image Copyright Protection Security Protocol in Internet Environment. *Chinese Journal of Computers,* 29:9(2006), 1722-1731.

[11] B.I. Federica, B. C.Giulia and N. Alessandro. Teaching multimedia data protection through an international online competition. *IEEE Transactions on Education*, 54:3(2011), 381-385.

[12] X.P.Zhang, S.Z. Wang. C. Chen. Asymmetric fingerprinting scheme without third party authorization. *Journal of Harbin Institute of Technology*, 38:7(2006), 727-730.