

Verifiable Text Watermarking Detection to Improve Security

Zhangjie Fu, Xingming Sun, Jiangang Shu, Lu Zhou and Jin Wang

School of Computer and Software & Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, 210044, China

*Email: wwwfzj@126.com, sunnudt@163.com, sjg.2008@qq.com,
zl_0713@163.com, wangjin@nuist.edu.cn*

Abstract

Digital watermarking technology plays an important role in the areas of copyright protection and identity tracing for owners of digital mediums. At present, the security of the watermarking scheme is facing a great threat. The security of a digital watermarking scheme must not depend on the scheme being kept secret. Zero knowledge-based watermark detection scheme (ZKWD) can achieve this aim. For ZKWD scheme, an owner can provide prove to a verifier that a digital medium in question indeed contains the owner's watermark information without revealing any secret key and watermark-related information. However, the existing ZKWD protocols are still facing some challenging problems, such as ambiguity attacks. In this paper, a public ZKWD protocol is proposed for plain text, and the homomorphic property of asymmetric encryption algorithm in the multiplication operation is used to prevent the owner from cheating by ambiguity attacks. Compared with existing methods, the security of our proposed ZKWD scheme is improved by using the improved feature extraction algorithm.

Keywords: *zero knowledge, watermark detection, ambiguity attacks, text watermarking*

1. Introduction

Text is the most widely used digital medium on the Internet, which brings people great convenience. But on the other hand, many problems are caused, such as copyright disputes and unauthorized copying etc. Digital watermarking has been recognized as one of most helpful technologies to protect the copyright of digital medium. So watermarking method's robustness is a critical issue, which can affect the practicability of the watermarking system. At present, the research about watermarking robustness has caused more attention. The traditional watermarking embedding and detection algorithms are public and the parameters (e.g., secret key) are confidential. This implies that the security of a watermarking system relies on the secret key rather than the watermarking embedding algorithm. At present, most of watermarking methods hold same secret key when embedding or detecting the watermarking information. This watermarking paradigm is called "symmetric" watermarking scheme.

However, the secret key in symmetric watermarking scheme may leak when detecting watermarking information because of the untrusted prover. The disclosure of the secret key can efficiently assist watermark-estimation attack [1] in removing watermarks. One solution to this problem is that we can use a secret key for watermarking embedding and a different but public key for watermarking detection. This is known as "asymmetric" watermarking scheme. Hartung and Girod [2] firstly proposed the idea of "asymmetric" watermarking.

However, the current asymmetric watermarking schemes only achieve limited robustness, and may suffer from security threat.

The security of a digital watermarking scheme must not depend on the scheme and secret key being kept secret. The secret key is inevitably revealed during the watermarking detection process. The untrusted prover could disclose the watermarking-related information when detecting the watermarking information. For attackers, furthermore, the leaked information is also sufficient to remove the watermarking from the disputed medium. This problem now is a common problem for all watermarking applications where the watermarking information has to be verified by an untrusted party.

In view of this security leakage, zero knowledge watermark detection (ZKWD) [3-7] scheme has been introduced by some researchers without obviously revealing the secret information, which can solve the problem of security effectively. The fundamental principle behind the zero knowledge watermark detection scheme is that a prover could convince a verifier that the prover certainly owns a secret key without revealing any watermarking-related information to verifier [8].

At present, the existing ZKWD schemes still suffer from some challenging problems, such as ambiguity attacks [9-11] *etc.*, one well known example for ambiguity attacks is that an adversary can create an ambiguous situation by deriving a forged watermark from a public work, and commits the forged watermarking information. Furthermore, the adversary is able to derive a watermark from existing non-watermarked medium in the public domain and claim ownership of them later.

In this paper, we propose a public zero knowledge watermark detection scheme which can prevent the owner from cheating by ambiguity attacks for plain text. Watermarking information is generated by using logistic chaotic mapping function from the extracted robust text features. Then the generated watermarking information is embedded into the text by using our proposed natural language information hiding method. Verifier can believe that the text in question indeed contains the owner's watermark during the watermark detection process through calculating the correlation between the watermark and the original text whose value is greater than a certain threshold.

The main contribution of this paper:

- 1) We are the first to present ZKWD method for plain text.
- 2) We produce a great correlation between the secret sequence (watermark) and the original text by using our improved feature extraction algorithm and watermarking generation method. This assists zero knowledge watermarking detection.
- 3) The homomorphic property of asymmetric encryption algorithm in the multiplication operation and the public parameters of the one-way function are used to prevent the owner from cheating by ambiguity attacks.

In the remainder of this paper, the following information is presented: In Section 2, related research is discussed, followed by the introduction of zero knowledge proof in Section 3. Then, the proposed zero knowledge watermark detection which includes four approaches is introduced in Section 4. In Section 5, security analysis is discussed. Finally, in Section 6, the paper concludes with some suggestions for future work.

2. Related Work

Zero knowledge watermark detection (ZKWD) scheme is a promising means to overcome the problem caused by disclosure of the secret key in the process of watermarking detection. ZKWD can be used to improve the security of digital watermarking scheme: it not only can conceal the required watermark information cryptographically, but also can prove the

presence of the hidden watermark information without revealing any secret information by an efficient zero knowledge proof system.

Craver, *et al.*, [4] firstly studied the ZKWD scheme and invertibility attacks, where an attacker can derive watermark information that is detectable in a given medium, and reverse the watermark embedding process. In their proposed protocol, the prover firstly generates many fake watermarks, then combines them with the legal watermark into one set, and then proves to verifier that there is a legal watermark which belongs to this set. For each round, a cheating prover can successfully pass the verification of the protocol with probability $1/2$. Nevertheless, if the prover and the verifier implement this protocol with polynomial iterations, then the cheating prover can pass the protocol only with negligibly low probability, *i.e.*, the verifier can be convinced with high probability.

Adelsbach, *et al.*, [5, 7, 9] proposed other zero knowledge watermark detection protocols, which use the zero knowledge proof as the sub-function of their scheme. For a watermarking scheme, the prover re-formulates the correlation between the original watermark and the extracted watermark into an appropriate form, then employs an existing zero knowledge protocol to prove for verifier that the watermark exists without leaking any secret information. Once again, their protocol, which is similar to Craver's protocol, needs to be implemented in a number of iterations.

Yu and Lu [12] presented a detector for zero knowledge watermark detection. Their scheme is resistant to ambiguity attacks by incorporating a one-way function based on the difficulty in finding Hamiltonian cycles in graphs. However, the security proof in their scheme is not sufficient, since the difficulty of inverting the one-way function does not imply the resistance against ambiguity attacks. Goldreich, *et al.*, [13] proposed that any conceivable property can be proven by an interactive zero knowledge proof protocol. Although the ZKWD protocol proposed by Goldreich, *et al.*, is constructive, their construction result becomes more complex.

As stated previously, the ZKWD protocol can be used to detect the presence of a watermark without revealing it, but the cheating behavior of a prover has not been avoided efficiently. A cheating prover could intentionally choose a "faked" watermark as though it were the legal watermark to pass the verification of the protocol and deceive the verifier. The existence of cheating prover in ZKWD can be regarded as a variant of ambiguity attacks.

Qiming Li, *et al.*, [16] proposed a zero knowledge watermark detector which can prevent the owner from cheating by ambiguity attacks for image. They used the cryptographically secure pseudo-random number generator and required that the watermarks are all generated from the generator with the same seed but with different indices associated with different works. However, similarly to Yu and Lu's method [12], their security is not sufficient to prevent ambiguity attacks, because there is not a great correlation between the secret sequence (watermark) and the original image. We also found that the ZKWD protocols presented so far are not established based on a robust watermarking scheme.

Therefore, motivated by the needs of sufficient security for watermarking, this paper proposes a public zero knowledge watermark detection protocol for plain text to prevent the owner from cheating by ambiguity attacks.

3. Zero Knowledge Proof

A zero knowledge proof of knowledge [13-14] is a two-party protocol between a prover and a verifier, which allows the prover to convince the verifier that he/she knows some secret information (proof of knowledge property), without that the verifier learns anything about them (zero knowledge property).

Zero Knowledge Cave [14] is a well-known example used to describe the main idea of zero knowledge proof. There are two parties in a zero knowledge proof protocol. The first party is known as a prover (Peggy) to prove the statement, while the second party is known as a verifier (Victor) to verify the statement.

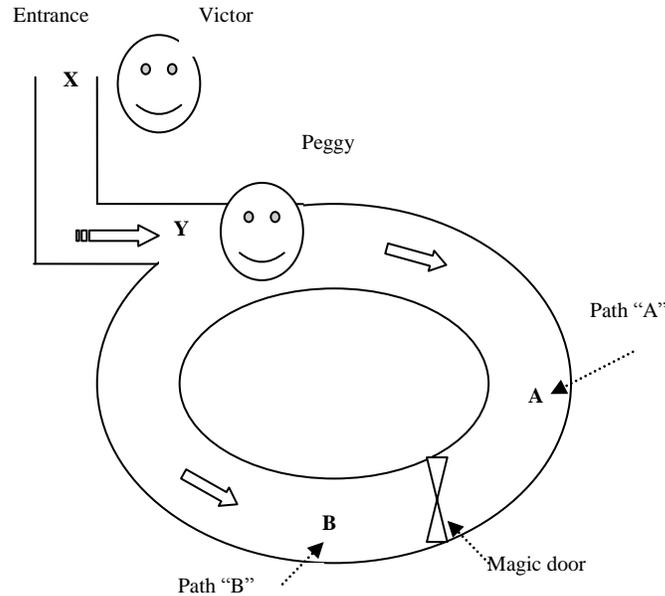


Figure 1. Zero Knowledge Cave

In this story, the circle cave has one entrance and a magic door which is placed inside the cave. The scenario depicted a proof protocol between Peggy and Victor, which help Peggy to prove her knowing the secret word which will open the magic door without revealing the secret word (which can open the door) to Victor. As shown by Figure 1 the cave paths are labeled as A for the left path and B for the right path. Both Victor and Peggy start from the cave entrance, X. First, Peggy enters the cave and randomly takes either path A or B while Victor must wait outside. Then, Victor will enter the cave to point Y and tell Peggy to appear from either path A or path B (randomly). Therefore, Peggy now can prove that she really knows the secret word by opening the magic door, if necessary and returns back to Y thru the path requested by Victor. For example, assume that Peggy knows the secret word and already she has gone inside the cave by path A and Victor ask her (randomly) to return back by path B, then she can open the magic door to appear on path A as requested by Victor. Assume Peggy does not know the secret word then this selection gives Peggy 50% chance of choosing properly. Repeating this protocol many times successfully makes Victor convinced that Peggy does actually know the secret word if Peggy can correctly appear all the time from the requested path specified by Victor.

Three main properties [5] of zero knowledge proof are showed as follows:

Completeness: The honest prover convinces the honest verifier that secret statement is true.

Soundness: Cheating prover can't convince the honest verifier that a statement is true (if the statement is really false).

Zero knowledge: Cheating verifier can't get anything other than prover's public data sent from the honest prover.

4. Zero Knowledge Watermark Detection

The zero knowledge watermark detection scheme proposed in this paper consists of watermarks generation, watermarking embedding and watermark detection. The basic idea of our proposed public ZKWD scheme is as follows:

Firstly, watermarks are generated by using logistic chaotic map from the robust text features which are extracted from plain text using our proposed feature extraction algorithm; then the watermarks are embedded into the plain text using our proposed natural language information hiding method (substitution of synonyms based on the semantic adjacent words). Therefore, there is a great correlation between the secret sequence (watermark) and the original text. When the zero knowledge watermark detection protocol is applied, verifier believes that the text in question contains the watermark claimed by prover by calculating the correlation, if the correlation value is greater than a certain threshold. The implementation process of the proposed method is illustrated in Figure 2.

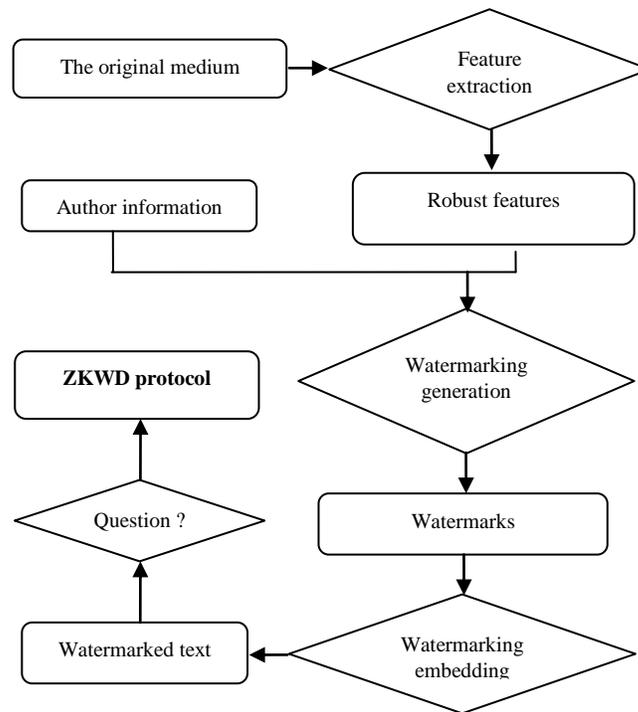


Figure 2. The Implementation Process of the Proposed Method

4.1 Watermarking Generation Method

In this section, a watermarking generation method based on logistic chaotic map is proposed. Chaotic systems have many important properties, such as the sensitive dependence on initial conditions and system parameters, the density of the set of all periodic points and topological transitivity, *etc.*, Most properties are related to some requirements such as mixing and diffusion in the sense of cryptography. Therefore, chaotic cryptosystems have more useful and practical applications.

The watermark information is generated from the extracted robust features and author information. The specific watermarking generation method is described as follows:

Input: the extracted robust features F , the author information A

Output: the watermarks W

- 1) Get the digital representation of the extracted robust feature using ASCII encoding and add it to F;
- 2) Get the digital representation of the author information using ASCII encoding and add it to F;
- 3) Scramble the sequence F using logistic chaotic map to S and keep the initial conditions and system parameters secret;
- 4) Get binarization sequence W (the watermarks) of the sequence S by defining the threshold;

According to the unidirectionality of chaotic sequence, it is very difficult for attackers to deduce the initial conditions and system parameters, because they are confidential and only the copyright holder owns them. Therefore the watermark generation function is a one-way function.

4.2 Watermarking Embedding Method

While zero knowledge watermark detection protocol is an important step towards secure watermark detection, it can only function reasonably if an embedded watermark can be detected/extracted from attacked media data. As a result, robustness against attacks is believed to be the prerequisite that must be satisfied before zero knowledge watermark detection protocol can be applied. Natural language watermarking is the most prominent scheme for text watermarking [15]. A robust watermarking embedding method is briefly proposed in this section. The generated watermarks will be embedded into the plain text by using our proposed natural language information hiding method that is substitution of synonyms based on the semantic adjacent words.

Firstly, the synonymy sets are created and classified. For the non-totally interchangeable synonymy sets, the context words are obtained from the semantic adjacent words by analyzing the dependency relationships, and then the synonym is selected with high probability of its cooccurrence of the semantic adjacent words. The method can effectively obtain the context words, and avoid the improper substitutions. Meantime, the method is able to avoid the usage of obscure words, and can resist efficiently the detection method based on substitution of synonyms.

The embedding secret key and extraction secret key are different in our proposed watermarking method, that is, we use a secret key for embedding and a different but public key for watermark detection, which is called asymmetric watermarking algorithm. This can avoid efficiently the problems which are caused by disclosure of the secret key, such as removing watermarks attack.

4.3 Public Zero Knowledge Watermark Detection

A cheating prover can intentionally choose a “faked” watermark as though it were the legal watermark to pass the verification of the protocol and deceive the verifier. Another case, the trusted third party is required in some zero knowledge watermark detection protocols to verify the signatures of the prover and verifier. However, the protocol could also be attacked by untrusted third party, and the involvement of trusted third party increases communication complexity. Therefore, in this section, a zero knowledge watermark detection protocol is proposed to prevent the owner from cheating by dishonesty prover and untrusted third party in the process of zero knowledge watermark detection.

According to the homomorphic property of asymmetric encryption algorithm in the multiplication operation and the public parameters of the hash function, anyone who can be considered as a verifier can detect that whether the data sent by prover is correct or not. By

calculating the linear correlation, verifier can believe that the text in question contains the watermark claimed by prover when the value of the correlation is greater than a certain threshold.

A cheating prover could, at best, succeed at cheating with probability $1/2$. If the protocol is repeated n times, the cheating prover could succeed at cheating with probability $1/2^n$. Again, this algorithm is performed until the probability of cheating falls beneath a certain threshold.

5. Security Analysis

In order to satisfy the requirements of security, four steps are introduced in this paper: the first one, called robust feature extraction, is to extract robust features from the plain text which will be considered as the seeds of watermark generation; the second step, watermark generation method, makes use of logistic chaotic map in order to generate watermarks to resist against the reversible attack; the third step is watermarking embedding algorithm by using the substitution of synonyms based on the semantic adjacent words to ensure the robustness of watermarking; the last one, asymmetric encryption algorithm, makes use of its homomorphic property to verify the correctness of data sent by prover.

These methods ensure that there is a great correlation between the watermark and the original text to prevent the owner from cheating by ambiguity attacks. An adversary can create an ambiguous situation by deriving a forged watermark from a published work and can achieve a certain threshold which is smaller than the threshold the paper used. Meantime, the protocol proposed in this paper does not need the trusted third party, which reduces the communication complexity and prevents the owner from cheating of the third party. Any verifier can check that whether the medium contains a watermark claimed by prover or not by using our proposed ZKWD protocol.

Compared with existing methods, the security of ZKWD is improved by producing a great correlation between the secret sequence (watermark) and the original text using the improved feature extraction algorithm. The homomorphic property of asymmetric encryption algorithm in the multiplication operation and the public parameters of the one-way function are used to prevent the owner from cheating by ambiguity attacks.

The detection protocol is computationally sound and satisfies the requirements of the three main properties of zero knowledge proof.

Completeness: The completeness requirement is easy to verify by inspection. The completeness of the whole protocol follows from the underlying detection, the homomorphic property of the asymmetric encryption algorithm and the completeness of the subproofs, what guarantees that an honest verifier will always accept a proof produced in an interaction with an honest prover.

Soundness: The soundness of the protocol comes from the soundness of the subproofs, that guarantees that the prover correctly produces the intermediate results, and the binding property between the watermarking and the original text, that assures that these results cannot be forged in a feasible time, because Prover can only cheat in ZKWD by cheating in the computation of hash function. However, for this Prover has to either breaks the soundness of one of the ZKWD protocol or the binding property of the one-way hash function which is assumed to be computationally infeasible.

Zero Knowledge: The zero knowledge property is also guaranteed by the zero knowledge of the sequentially composed subproofs and the statistically hiding property of the used asymmetric encryption algorithm. A simulator can be built that, given the random choices of the verifier, can produce an indistinguishable output of an accepting protocol, just using the existing simulators for the zero knowledge subproofs, when these are only known by the

prover. So the detection is performed without providing to verifier any information additional to the presence of the watermark.

6. Conclusions

A difficult problem in traditional watermarking scheme is that the security of watermark needs to be enhanced. The secret key is inevitably revealed during the watermarking detection process. The untrusted prover could disclose the watermarking-related information when detecting the watermarking information. ZKWD protocol can be used to improve the security of watermark scheme, and it can achieve the aim that the leaked watermark-related knowledge is zero.

In this paper, a public zero knowledge watermark detection protocol is proposed to prevent the owner from cheating by ambiguity attacks. Four steps are concluded in our proposed zero knowledge watermark detection scheme to achieve this goal: robust feature extraction method, watermarks generation method, watermarking embedding method and zero knowledge watermark detection protocol. Any verifier can check that whether the medium contains a watermark claimed by prover or not. The proposed method can satisfy the three requirements of zero knowledge proof of identity: completeness, soundness, zero knowledge. Meantime, the method ensures the security for watermark verification in the watermarking detection process without revealing any secret information related to watermarking.

Future work is to study a non-interactive zero knowledge watermark detector robust to sensitivity attacks.

Acknowledgements

This paper is a revised and expanded version of a paper entitled "Plain Text Zero Knowledge Watermarking Detection Based on Asymmetric Encryption" presented at CIA 2014, Angeles City (Clark), Philippines, April 24 -26, 2014. This work is supported by the NSFC (61232016, 61173141, 61173142, 61173136, 61103215, 61373132, 61373133), GYHY201206033, 201301030, 2013DFG12860, BC2013012, PAPD fund, Hunan province science and technology plan project fund (2012GK3120), the Scientific Research Fund of Hunan Provincial Education Department (10C0944), and the Prospective Research Project on Future Networks of Jiangsu Future Networks Innovation Institute (BY2013095-4-10).

References

- [1] C. S. Lu and C. Y. Hsu, "Content-Dependent Anti-Disclosure Image Watermark", Proceedings of 2nd Int. Workshop on Digital Watermarking, LNCS 2939, Seoul, Korea, (2003), pp. 61-76.
- [2] F. Hartung and B. Girod, "Fast Public-Key Watermarking of Compressed Video", Proc. IEEE Int. Conf. on Image Processing, Santa Barbara, CA, USA, (1997) October, pp. 528-531.
- [3] K. Hirotsugu, "An image digital signature system with ZKIP for the graph isomorphism problem", Proceedings of the 3rd IEEE Conference on Image Processing, Piscataway, NJ, (1996), pp. 247-250.
- [4] S. Craver, "Zero knowledge watermark detection", The 3rd Int' Workshop Information Hiding. LNCS 1768, Berlin: Springer-Verlag, (2000), pp. 101-116.
- [5] A. Adelsbach and Sadeghi R, "Zero knowledge watermark detection and proof of ownership", The 4th Int' Workshop Information Hiding. LNCS2137, Berlin: Springer Verlag, (2001), pp. 273-287.
- [6] K. Gopalakrishnan, N. Memon, P. Vora, "Protocols for watermark verification: Multimedia and security", IEEE Multimedia, vol. 8, no. 4, (2001), pp. 66-70.
- [7] A. Adelsbach, S. Katzenbeisser and A. R. Sadeghi, "Cryptography meets watermarking: Detecting watermarks with minimal or zero knowledge disclosure", In Proc. of the European Signal Processing Conf. (EUSIPCO 2002), (2002), pp. 446-449.

- [8] X. X. Zou, Q. Dai, C. Huang and J. T. Li, "Zero-Knowledge watermark verification protocols", *Journal of Software*, vol. 14, no. 9, (2003), pp. 1645-1651.
- [9] A. Adelsbach, S. Katzenbeisser and A.R. Sadeghi, "Watermark detection with zero-knowledge disclosure", *Multimedia Systems*, Springer-Verlag, (2003), pp. 266-278.
- [10] J. R. Troncoso-Pastoriza and F. Perez-Gonzalez, "Zero-knowledge watermark detector robust to sensitivity attacks", in 8th ACM Multimedia and Security Workshop, Geneva, Switzerland, (2006) September, pp. 97-107.
- [11] H. Zhang, Z. Yuan and Q. Wen, "A digital signature schemes without using one-way hash and message redundancy and its application on key agreement", 2007 IFIP International Conference on Network and Parallel Computing Workshops, Washington, DC, USA, (2007), pp. 873-878.
- [12] C.-M. Yu, and C.-S. Lu, "Robust non-interactive zero-knowledge watermarking scheme against cheating prover", In 7th ACM Multimedia and Security Workshop, New York City, NY, USA, (2005), pp. 32-41.
- [13] O. Goldreich, S. Micali and A. Wigderson, "Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems", *Journal of the ACM*, vol. 38, no. 3, (1991), pp. 690-728.
- [14] I. Damgard, "Commitment schemes and zero-knowledge protocols", In *Lectures on data security: modern cryptology in theory and practise*, Lecture Notes in Computer Science, Springer-Verlag, Berlin Germany, vol. 1561, (1998), pp. 63-86.
- [15] Z. Liu, X. Sun, Y. Liu, L. Yang, Z. Fu, Z. Xia and W. Liang, "Invertible Transform-Based Reversible Text Watermarking", *Information Technology Journal*, vol. 9, no.6, (2010), pp. 1190-1195.
- [16] Q. Li and E. C. Chang, "ZeroKnowledge Watermark Detection Resistant to Ambiguity Attacks", *Proceedings of the 8th workshop on Multimedia and security*, (2006), pp. 493-498.
- [17] Z. Fu, X. Sun, N. Linge and L. Zhou, "Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query", *IEEE Transactions Consumer Electronics*, vol. 60, no.1, (2014), pp. 164-172.
- [18] Z. Fu, X. Sun, Z. Xia, L. Zhou and J. Shu, "Multi-keyword Ranked Search Supporting Synonym Query over Encrypted Data in Cloud Computing", *IEEE 32nd International Performance Computing and Communications Conference (IPCCC 2013)*, San Diego, CA, (2013), pp. 1-8.
- [19] Z. Fu, X. Sun and Y. Liu, "Text Split-based Steganography in OOXML Format Documents for Covert Communication", *Security and Communication Networks*, vol. 5, no.9, (2012), pp. 957-968.
- [20] A. Rezai and P. Keshavarzi, "A New Left-to-Right Scalar Multiplication Algorithm Using a New Recoding Technique", *International Journal of Security & Its Applications*, vol. 8, no. 3, (2014), pp. 31-38.
- [21] N. Thirananant and H. J. Lee, "A Design of e-Healthcare Authentication Framework with QR Code", *International Journal of Security & Its Applications*, vol. 8, no. 3, (2014), pp. 79-86.

Authors



Zhangjie Fu, he received his BS in education technology from Xinyang Normal University, China, in 2006; received his MS in education technology from the College of Physics and Microelectronics Science, Hunan University, China, in 2008; obtained his PhD in computer science from the College of Computer, Hunan University, China, in 2012. Currently, he works as an assistant professor in School of Computer and Software, Nanjing University of Information Science and Technology, China. His research interests include cloud computing, digital forensics, network and information security.



Xingming Sun, he received his BS in mathematics from Hunan Normal University, China, in 1984; his MS in computing science from Dalian University of Science and Technology, China, in 1988; and his PhD in computing science from Fudan University, China, in 2001. He is currently a professor at the College of Computer and Software, Nanjing University of Information Science and Technology, China. In 2006, he visited the University College London, UK; he was a visiting professor in University of Warwick, UK, between 2008 and 2010. His research

interests include network and information security, database security, and natural language processing.



Jiangan Shu, he received his BE in Network Technology and Engineering from Nanjing University of Information Science & Technology (NUIST), Nanjing, China, in 2012. He is currently pursuing his MS in computer science and technology at the School of Computer & Software, Nanjing University of Information Science and Technology, China. His research interests include cloud security, steganography, network and information security.



Lu Zhou, she received her BE in Software Engineering from Nanjing University of Information Science and Technology, China, in 2012. She is currently pursuing her MS in computer science and technology at the School Of Computer and Software, Nanjing University of Information Science and Technology, China. Her research interests include network and information security, steganography, digital watermarking, copyright protection technology.