Text Coverless Information Hiding Method Based on Hybrid Tags

Yulei Wu[1,2,3], Xingming Sun[1,2,3]

[1] Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, China

[2] Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, China

[3] School of Computer and Software, Nanjing University of Information Science & Technology, China

wyl940314@hotmail.com, sunnudt@163.com

Abstract

In recent years, a novel information hiding method called coverless information hiding, which has the anti-detection ability due to its non-modification of the stego cover, has drawn the attention of researchers. In coverless information hiding, the communication party uses tags to locate secret messages instead of sharing encryption key so as to prevent the third party from extracting them. This paper proposes a new method of coverless information hiding based on two tag selecting strategies, using two sub-protocols to determine the tags by the identity of the receiver. Experimental results prove that the coverless information hiding is able to resist the steganalysis. Besides, the results reveal that the proposed method in this paper has a higher success rate and hiding capacity than these methods based on single tag selection.

Keywords: Information hiding, Chinese character encoding, Chinese mathematical expression, Steganography, Coverless information hiding

1 Introduction

Entering the information age, the speed of information transmission keeps increasing owing to the constant improvement of internet technologies. As a result, information security draws more and more attention from researchers, including the digital forensics [1-2], copyright protection [3-4], information hiding and so on. Information hiding is intended to take advantages of the insensitivity of human senses and the redundancy of the digital signal so as to conceal the secret message into the public information. This kind of public information with secret message is called stego cover. According to the type of the stego cover, the information hiding can be divided into text based and multimedia based steganography [5-6]. There are several ways to generate the stego cover with embedded messages in conventional information hiding.

The most common way is to embed hiding message by modifying the original carrier. The method proposed in [7] is to embed secret messages by changing the space, height and width of the characters. Some other researchers do this by modifying in properties without influencing the visual effect [8]. Wu and Liu use the combination of the text and binary image to embed hiding information [9]. In [10], based on embedding technology and leaving local and global semanteme unchanged, the authors modify the text with different efforts to conceal the information.

There are also some information hiding methods embed the secret messages by text production. An information hiding method which produces text based on probability statistics is proposed in [11]. Although the text produced by this method corresponds to the statistic features of natural languages, it is hard to escape from the identification of human eyes.

Conventional researches for information hiding have achieved various success. Nevertheless an unavoidable problem exists: the security of information hiding is hard to be guaranteed. There is no way to resist steganalysis by computers [12-13] or human eyes since all of the stego cover applied in these information hiding methods are not natural.

A new information hiding method called coverless information hiding is proposed in recent years. “Coverless information hiding” is an information hiding technology which retrieves natural carriers with features as the stego cover. It needs no access to any other carriers and can generate stago cover directly regarding secret information. The authors proposed an information hiding method, one of which is based on text big data in [14-17] and the other one is based on images in [18-20]. There are two major differences between coverless information hiding and conventional

*Corresponding Author: Xingming Sun; E-mail: sunnudt@163.com
DOI: 10.3966/160792642018051903003

information hiding: Firstly, coverless Information hiding leaves the carrier unmodified and transmits the hiding information through built-in attributes of the stego cover directly. Secondly, for coverless information hiding, no additional information is required during the transmission process. The receiver only requires the prearranged protocol and the stego cover to extract the hidden information. Coverless information hiding has the ability to resist the existing steganalysis since it has the two features mentioned above.

The remainder of this paper is organized as follow: the related work will be introduced in section 2. Then the new method of tag selection of coverless information hiding proposed in the paper and the algorithm process will be discussed in section 3. Afterwards, section 4 introduces a series of experiments. Finally, section 5 will give a summary and an outlook.

2 Related Work

Unlike the conventional information hiding methods, coverless information hiding retrieves natural texts with features to transmit secret message. There're many methods to retrieve information in encrypted domain [21-22]. Besides, the sender doesn't transmit any assistant information to the receiver. Thus, the challenge of this algorithm is the extraction of the secret message and the retrieve of the stego cover. The accepted way to extract the message is using tags in front of the message as a locator, and the method to accelerate retrieve is building index.

2.1 Information Hiding Methods Based on Tag Locating

In text coverless information hiding, stego cover transmitted is actually a natural and public text. The receiver cannot extract the hidden information using the method in conventional information hiding which is extracting features. In order to provide a certain location of hiding, tags are utilized in front of the ciphertext to remark the location of it.

Every time when transmitting the information, the sender and the receiver share a set of tags. The tag is one or several Chinese characters that contain certain features. For example, the UNICODE of the character and the unique property of the Chinese character can be utilized as features. After getting the stego cover, the receiver scans the text and marks all the tags used in the transmission to get the accurate location of the hidden information and extract the secret message.

2.2 Inverted Index

Inverted index originates from the practical application needs of finding records based on the value of the property. Unlike other index technologies to determine the value of the property according to the records, inverted index determines the location of the records by the attribute values. The creation of the inverted index is a process of transferring the texts into an inverted list table. The inverted list consists of many records of keywords. Every row of the table records the files which includes this keyword (DocId), the frequency of the keyword (TF) and other information. When searching for the files that includes certain keywords, the user only needs to seek the inverted table of this keyword.

3 The Proposed Method

The fundamental principle of coverless information hiding is to retrieve texts. By means of searching some documents with certain features, the sender can proceed the information hiding without modifying the cover itself. The stego cover here is actually unmodified as natural texts. With the help of the prearranged protocols, the sender and the receiver can communicate with each other secretly. In this paper, a new method of coverless information hiding based on tag locating is proposed. Details and the core parts of this algorithm will be discussed as follows.

3.1 Tag Design

Tag is used to provide the location of the ciphertext. The communication party only transmits the natural texts as the stego cover in coverless information hiding. Due to this limitation, the extraction of the ciphertext will be impossible without the help of the tag. Tag is a part of the stego cover. Besides, the transmitted texts cannot be modified. Thus, the tag is actually some unmodified characters with certain features. However, not all of the Chinese characters can be utilized as tags. So the tag selection is the core part of the algorithm. All tags selected must satisfy the randomness and the universal to transmit different ciphertext using different texts. While the original Chinese character doesn't meet the requirement above due to the limited property of Chinese. Therefore, the conversion of the Chinese characters is required. In this paper, two methods of tag selection based on Chinese conversion, the Chinese character encoding and the Chinese mathematical expression, are proposed.

The tag selection based on the Chinese character encoding is to translate several straight Chinese characters into a binary stream. Firstly, the character is translated into UNICODE. Then, the odevity of the UNICODE is considered to decide the value 1 or 0. 6-bit binary stream is utilized as a tag. The word or phrase corresponding to the tag is the secret information. 6-bit binary stream provides the randomness and the universal to a certain extent.

The tag selection based on mathematical expression is to turn the Chinese characters into the Chinese

mathematical expressions [23]. The numerical value in the Chinese mathematical expression represents different components of the characters. Different value denotes different tags. Like the method above, the phrase or the word at the back of the character which contains certain tag is a hidden information. The Chinese character itself doesn't have the universality. Therefore, the Chinese characters are split into smaller units to overcome this problem.

3.2 Tag Protocol

The significance of tag is introduced above. To provide the security, different tags are used in different communication. It is the protocol that decides which tags to be used in the communication each time. Different tag sets will be generated through different identities of the communicating parties to make the transmission secure. The proposed method here combines the two ways of tag selection mentioned above in the protocol. In this paper, a simplified protocol prototype is designed.

The protocols are as follows: firstly a specific ID= $I_1 I_2 ... I_n$ is set for each receiver (n represents the total length of the ID, I_i is an integer ranged from 0 to 9). For each specific ID, there are two sub protocols of tag selection.

The first one is the sub-protocol based on Chinese encoding: In this protocol, 2^6=64 tags (000000~111111) are utilized. The tag set $Q = \{q_1, q_2, ...q_n\}$ is obtained based on receiver's ID (the sender and the receiver share this tag set, $q_i = q_1 + i * S$, if $q_i > q_{\max}$, ($q_{\max}$ =63), let q_i = mod(q_i,63) , q_i and S are determined by the user's ID. In this paper, q_1 is the last but one digit of the ID, while S is the last digit of the ID for experiments).

The second one is the protocol based on Chinese mathematical expression: 500 components of the Chinese mathematical expression are sifted through and 150 of them with the highest frequency are taken as tags. These tags are sorted with number 1~150. The tag set $W = \{w_1, w_2, ..., w_n\}$ is obtained based on receiver's ID ($w_i = w_1 + i * S$, if $w_i > w_{\max}$ ($w_{\max}$ =150), let $w_i = \text{mod}(w_i, 150)$, w_i is determined by the user's ID as the former sub-protocol).

Example: Suppose ID=75, q_1 =7, w_1 =9, S=5, Q ={7,12, 17...}, W ={9,14,19...}.

3.3 Information Hiding

In conventional information hiding, the process of information hiding is generating artificial texts or modifying natural texts. However, in coverless information hiding methods, the process as mentioned above, is more like retrieving. The specific procedure is discussed below.

In the hiding phase, the secret information is segmented into several keywords, with the help of Chinese segment algorithm. Then the tag sets are obtained according to the tag protocol and the receiver's ID. Each keyword and each tag is a one-to-one match. Afterwards, for each keyword, retrieve the texts that has certain features (the word following the first appearance of the tag in this text is the keyword) and store them into sets. Finally, from each set, one text that meets certain rule will be selected (e.g. The $\boldsymbol{R_T}$ of the latter text selected must be greater than the $\boldsymbol{R_T}$ of the former text selected. $\boldsymbol{R_T}$ is an attribute or feature of this text.), so that the receiver can extract the keyword in a right order. And all the selected texts are the stego cover. The pseudocode is shown below in Figure 1.

Input: Receiver's ID, Secret information
Output: Stego cover texts

Begin:
(1) Segment the secret information, get $M = \{m_1, ...m_k\}$;
(2) Obtain the tag sets Q , W ;
(3) For each keywords m_i : (i=1,2,...)
(4) Retrieve the texts that include the combination $q_i + m_i$, and put them into set d_i ;
(5) Retrieve the text that include the combination $w_i + m_i$, and put them into set e_i ;
(6) Get the union $D_i = d_i \cup e_i$;
(7) Set $\boldsymbol{J}$ = 0;
(8) For each set D_i :
(9) Pick one text that meet the rule($\boldsymbol{R_T > J}$);
(10) Let $\boldsymbol{J} = \boldsymbol{R_T}$;
(11) All the selected texts are the stego cover;
End.

Figure 1. Pseudocode of information hiding

3.4 Index Creating

The core of the proposed method is retrieving. For the purpose of raising efficiency of retrieving files, index of the database is necessary. Different from conventional index, the feature of every keyword needs to be recorded in the proposed method. Therefore, inverted index method is utilized for the texts and the index result is saved into the database. The index demands to record the keyword and the tag in front of the keyword, as well as the document path and the attribute $\boldsymbol{R_T}$.

The process of building an index is as follows:

Firstly, hash algorithm is applied to the first several Chinese characters of each text. The result will be taken as $\boldsymbol{R_T}$, the basis of order when extracting the information eventually.

Secondly, the 4 characters following the first appearance of each tag in every text is chosen to be a sub-string. This sub-string is segmented by maximum match segmentation method (Four is chosen as the length of division because 90% of the Chinese words are in a four characters limit). The first character or word after division is chosen. For example: In some text, the four characters after some tag are "今天天气", the result after division should be "今天|天气". Then "今天" is chosen as the keywords following this tag.

Finally, drawing data tables for each tag under these two sub-protocol. The first column for each table is the character or word located in each text by each tag. The second column is absolute path and the feature $\boldsymbol{R_T}$ of any text meets this condition. And the separator ";" is used for distinction.

Example: Suppose the $\boldsymbol{R_T}$ of the text is R, the keyword after tag "000000" (Chinese encoding based protocol) is K1, and the keyword after the tag "24" (Chinese mathematical expression based protocol) is K2. Then {K1(keyword value), text path(R)(path value)} will be inserted into data table tag000000, and {K2(keyword value), text path(R)(path value)} will be inserted into data table tag24.

3.5 Information Extraction

Information extraction is the inverse process of hiding. In information extraction phase, the receiver obtains the tag sets according to the protocol and the ID after receiving stego cover.

To avoid the disorder of the extracted information, the stego cover texts will be sorted by the attribute $\boldsymbol{R_T}$ then. Due to the combination of two tag selection methods, the receiver can not tell which tag is used in the stego cover. Therefore, while extracting the keyword, the receiver needs to use two tags from the two tag sets in every position. Thus there are two possible keywords in each position, which leads to 2^k different possible combination of keywords (k is the total amount of keywords). Then, semantic analysis is utilized here to reduce the possibilities. Afterwards, the confidence level of each possibility is computed. The receiver will get the secret information by manual selection.

The pseudocode is shown in the Figure 2.

4 Experimental Results

To test the performance of the proposed method, various kinds of experiments are implemented. The test sets of secret information are obtained from the Sougou Lab. Sougou Lab is an open experimental platform. This paper uses the web data compilation as test sets, the composition of the test sets is shown below in Table 1.

Input: Receiver's ID, Stego cover tests
Output: Secret information

Begin:
(1) Obtain the tag sets Q , W ;
Sort the stego cover texts by $\boldsymbol{R_T}$ in ascending Order;
(3) For each text:
(4) Extract the phrase/word following q_i in this text;
(5) Extract the phrase/word following w_i in this text;
(6) Combine the possible keyword, and we will get 2^k possible combination(k is the amount of keywords);
(7) Get rid of the combination which contains semantic error;
(8) Possibility ranking;
(9) Manual selection;
End.

Figure 2. Pseudocode of the information extraction

Table 1. Data sets

Secret information category	Total size (MB)	Quantity	Average size(KB)
Education	6.8	1990	3.50
Military	5.0	1990	2.57
Traveling	4.2	1990	2.16
Society	6.7	1990	3.45
Sports	3.2	1990	1.65
culture	11.4	1990	5.87

4.1 Anti-detection Ability

The major advantages of the coverless information hiding over the conventional one is its ability to resist the steganalysis. To test the anti-detection ability of the proposed method, this paper compares different sizes of stego cover texts generated by this method, using the steganalysis method proposed in [24], the experimental results are shown below.

The conventional information hiding stego cover set consists of 450 txt files (342 natural texts randomly picked and 108 texts embedded with secret information by synonym substitution steganography (SSS, for short)). The other three sets all consist of 450 txt files picked in the local database. Thus, these texts by definition are stego cover texts in coverless information hiding. False negative rate is the proportion of the stego covers which escape the steganalysis in all stego covers. False positive rate is the proportion of the mistakenly recognized normal texts in all normal texts. The results are shown in Table 2.

Table 2. Steganalysis resuts

Texts	Total amount	False negative rate(FN)	False positive rate(FP)
1kb sized	450	99.1%	-
2kb sized	450	99.3%	-
5kb sized	450	98.2%	-
8kb sized	450	98.2%	-
SSS	450	5.5%	0.8%

The experimental results indicate that the proposed method can resist the detection of the steganalysis method due to the non-modification of the stego cover. The conventional information hiding methods modify the stego cover to embed secret information, which leads to irresistibility of statistical based steganalysis.

As is shown in the Table 2, the FP in conventional information hiding plus the FN in coverless information hiding is approximately equal to 100%. That is to say, the texts marked as abnormal in the first 4 rows of the table are mistakenly recognized by the steganalysis. Although some of the stego cover texts in coverless information hiding is recognized as abnormal texts mistakenly, the false negative rate of the coverless information hiding method is still much higher than that of the conventional ones.

4.2 Success rate of Hiding

Success rate of hiding is an important indicator of the algorithm's performance. The principle of the coverless steganography causes that the hiding of the ciphertext is not always successful. If there are no texts that meet the requirements, the information hiding will fail.

First, this paper compares the success rate of information hiding using different tag selection methods. Four hundred and fifty 1kb sized, 2kb sized, 5kb sized and 8kb sized texts from the secret information sets are used in the experiments.

The success rate $\Re$ is defined as below.

$$\Re = \frac{C}{T} \times 100\% \quad (1)$$

Where C denotes the number of successfully hidden texts and T means the total number of the test sets. In this paper, the results are shown in Table 3.

Table 3. The success rate of hiding with the different tag selection methods

Tag selection method		Chinese character encoding	Chinese mathematical expression	The proposed method
Test set	1kb	96.9%	97.7%	98.9%
	2kb	99.1%	98.9%	100%
	5kb	96.9%	97.6%	98.9%
	8kb	96%	98.4%	99.6%

It is clearly shown in the experimental results that the proposed method combines the advantages of the first 2 tag selection methods. Theoretically, the success rate of the proposed method is greater than or equal to the first 2 methods. The average success rate of the proposed method is around 99.3%. When the size of ciphertext is around 2 kb, the success rate of the information hiding obtains its peak value.

The success rate of different text database is tested as well by using 50 secret information texts from category culture, society and sports. The results are shown in Table 4.

Table 4. The success rate of hiding with the different text databases

Stego cover database		Novels	News	Classical literature
Secret information category	culture	98%	96%	98%
	society	94%	98%	96%
	sports	94%	94%	92%

It is indicated in the results that the category of the stego cover has a certain degree of influence on the success rate. If the category of the secret information is the same as or similar to the category of the stego cover, the success rate will be relatively higher. While generating the stego cover, the identity and the dabbling field of the receiver ought to be considered so as to enhance the success rate and the security of the information hiding.

4.3 Extraction Analysis

Extraction difficulty is another factor that shows the performance. As mentioned above, if the amount of the keywords after segmentation is k, there will be 2^k possible secret messages. Such huge amount of possibilities will make manual selection difficult. Therefore, after combination of the keywords, those combinations with semantic errors are got rid of. Secret messages with different length (different k) are used to test the ratio of the messages with errors to all possibilities (Error ratio).

The experimental results are shown in Table 5.

Table 5. The ratio of the possibilities with semantic errors

Amount of the keywords	All possibilities	Error ratio
2	4	71.25%
3	8	83.75%
4	16	91.88%
5	32	96.25%
6	64	98.13%
7	128	98.59%
8	256	99.37%

According to the results, most of the possibilities contain semantic errors. The error ratio achieves stability when k=7 or higher. Although the amount of possible messages increases exponentially, most of the possibilities are not idiomatic. As a result, the actual amount of messages after removing the error messages is very small, the average amount is around 1.5. Therefore, the extraction difficulty is relatively low and the success rate of extraction is acceptable. With the help of the possibility ranking, the success rate of the extraction will be further improved.

4.4 Hiding Capacity

In this paper, only the first appearance of each tag is used in order to simplify the calculation. Therefore, the average size of the stego text cannot be utilized as a parameter when calculating the hiding capacity. The hiding capacity in this paper is defined below as V_i.

$$V_i = \frac{h}{H} \text{(characters per text)} \quad \textbf{(2)}$$

Where i is the serial number of the tested text, h means the amount of Chinese characters to be hidden in the stego cover, and H refers to the total amount of the stego cover texts.

To compare the hiding capacity of different tag selection methods, 50 texts (around 2kb) from all categories of the secret information database are used. The results are shown below in Figure 3.

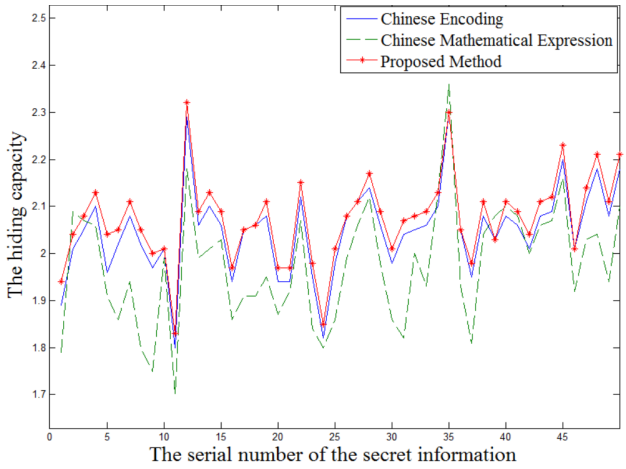


Figure 3. Hiding capacity

In the experimental results, it is clear that different tag selection methods lead to different hiding capacity. The average capacity of the Chinese mathematical expression based method is about 1.97, while the average capacity of the Chinese encoding based method is 2.03. The proposed method improve the capacity by 2-5 percent, reaching 2.07.

5 Conclusion

This paper proposes a coverless information hiding method based on hybrid tags, using two ways to select tags. Unlike the conventional methods, the proposed method transmits natural texts to carry the ciphertext. The sender retrieve the texts with certain features, while the receiver extract the information with the help of the tags. The experimental results show that the proposed method has the ability to resist steganalysis.

Compared to the conventional methods, the proposed method has a lower hiding capacity. However, it provides a way of improving the capacity for the coverless steganography researches. This paper uses the first appearance of each tag, making the utilization rate of texts low. Improving the utilization rate is another focus of our future work.

Acknowledgements

This work is supported by the National Key R&D Program of China under grant 2018YFB1003205; by the National Natural Science Foundation of China under grant U1536206, U1405254, 61772283, 61602253, 61672294, 61502242; by the Jiangsu Basic Research Programs-Natural Science Foundation under grant numbers BK20150925 and BK20151530; by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund; by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China.

References

[1] J. Li, X. Li, B. Yang, X. Sun, Segmentation-Based Image Copy-Move Forgery Detection Scheme, *IEEE Transactions on Information Forensics & Security*, Vol. 10, No. 3, pp. 507-518, March, 2015.

[2] C. Yuan, X. Li, Q. Wu, J. Li, X. Sun, Fingerprint Liveness Detection from Different Fingerprint Materials Using Convolutional Neural Network and Principal Component Analysis. *Computers, Materials & Continua*, Vol. 53, No. 3, pp. 357-371, September, 2017.

[3] Z. Zhou, C.-N. Yang, B. Chen, X. Sun, Q. Liu, Q. M. Wu, Effective and Efficient Image Copy Detection with Resistance to Arbitrary Rotation, *IEICE Transactions on Information and Systems*, Vol. E99-D, No. 6, pp. 1531-1540, January, 2016.

[4] Z. Zhou, Y. Wang, Q. M. J. Wu, C.-N. Yang, X. Sun, Effective and Efficient Global Context Verification for Image Copy Detection, *IEEE Transactions on Information Forensics and Security*, Vol. 12 No. 1, pp 48-63, January, 2017.

[5] W.-C. Hsieh, N.-I. Wu, C.-M. Wang, A Novel Data Hiding Scheme for Binary Images with Low Distortion, *Journal of Internet Technology*, Vol. 11, No. 7, pp. 1057-1069, December, 2010.

[6] Z. Wang, Q. Mao, C. Chang, Q. Wu, J. Li, A Data Lossless Message Hiding Scheme without Extra Information, *Journal of Internet Technology*, Vol. 15, No. 4, pp. 657-669, July, 2014.

[7] J. T. Brassil, S. Low, N. Maxemchuk, Copyright Protection for the Electronic Distribution of Text Documents, *Proceedings of the IEEE*, Vol. 87, No. 7, pp. 1181-1196, July, 1999.

[8] R. Meng, S. Rice, J. Wang , X. Sun, A Fusion Steganographic Algorithm Based on Faster R-CNN, *Computers Materials & Continua*, Vol. 55, No. 1, pp. 001-016, March, 2018.

[9] M. Wu, B. Liu, Data Hiding in Binary Images for Authentication and Annotation, *IEEE Transactions on Multimedia*, Vol. 6, No. 4, pp. 528-538, August, 2004.

[10] M. L. Mali, N. N. Patil, J. B. Patil, Implementation of Text Watermarking Technique Using Natural Language Watermarks, *IEEE 2013 International Conference on Communication Systems and Network Technologies*, Gwalior, India, 2013, pp. 482-486.

[11] P. Wayner, Mimic Functions, *Cryptologia*, Vol. 16, No. 3, pp. 193-214, July, 1992.

[12] Z. Xia, X. Wang, X. Sun, Q. Liu, N. Xiong, Steganalysis of LSB Matching Using Differences between Nonadjacent Pixels, *Multimedia Tools and Applications,* Vol. 75, No. 4, pp. 1947-1962, February, 2016.

[13] Z. Xia, X. Wang, X. Sun, B. Wang, Steganalysis of Least Significant Bit Matching Using Multi-order Differences, *Security & Communication Networks*, Vol. 7, No. 8, pp. 1283-1291, August, 2014.

[14] Z. Zhou, Y. Mu, C.-N. Yang, N. Zhao, Coverless Multi-keywords Information Hiding Method Based on Text, *International Journal of Security and Its Applications*, Vol. 10, No. 9, pp. 309-320, September, 2016.

[15] Z.-L. Zhou, Y. Cao, X.-M. Sun, Coverless Information Hiding Based on Bag-of-Words Model of Image, *Journal of Applied Sciences*, Vol. 34, No. 5, pp. 527-536, September, 2016.

[16] H. Sun, R. Grishman, Y. Wang, Active Learning Based Named Entity Recognition and Its Application in Natural Language Coverless Information Hiding, *Journal of Internet Technology*, Vol. 18, No. 2, pp. 443-451, March, 2017.

[17] X. Chen, H. Sun, Y. Tobe, Z. Zhou, X. Sun, Coverless Information Hiding Method Based on the Chinese Mathematical Expression, *First International Conference on Cloud Computing and Security*, Nanjing, China, 2015, pp. 133-143.

[18] Z. Zhou, H. Sun, R. Harit, X. Chen, X. Sun, Coverless Image Steganography Without Embedding, *First International Conference on Cloud Computing and Security*, Nanjing, China, 2015, pp. 123-132.

[19] Y. Cao, Z. Zhou, X. Sun, C. Gao, Coverless Information Hiding Based on the Molecular Structure Images of Material, *Computers Materials & Continua*, Vol. 54, No. 2, June, pp. 197-207, 2018.

[20] C. Yuan, Z. Xia, X. Sun, Coverless Image Steganography Based on SIFT and BOF, *Journal of Internet Technology*, Vol. 18, No. 2, pp. 435-442, March, 2017.

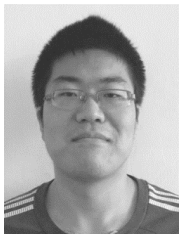
[21] Z. Fu, X. Wu, C. Guan, X. Sun, K. Ren, Toward Efficient Multi-keyword Fuzzy Search over Encrypted Outsourced Data with Accuracy Improvement, *IEEE Transactions on Information Forensics and Security*, Vol. 11, No. 12, pp. 2706-2716, December, 2016.

[22] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, K. Ren, A Privacy-preserving and Copy-deterrence Content-based Image Retrieval Scheme in Cloud Computing, *IEEE Transactions on Information Forensics and Security*, Vol. 11, No. 11, pp. 2594-2608, November, 2016.

[23] X. Sun, J. Yin, H. Chen, Q. Wu, X. Jing, On Mathematical Expression of a Chinese Character, *Journal of Computer Research and Development*, Vol. 39, No. 6, pp. 707-711, June, 2002.

[24] Z. Chen, L. Huang, W. Yang, Detection of Substitution-based Linguistic Steganography by Relative Frequency Analysis, *Digital Investigation*, Vol. 8, No. 1, pp. 68-77, July, 2011.

Biographies



Yulei Wu received the B.E. degree in software engineering from the Nanjing University of Information Science & Technology in 2016, China. He is currently working towards the MS degree in computer science and technology at the College of Computer and Software, in Nanjing University of Information Science & Technology, China. His research interest includes network and information security.

Xingming Sun is a professor in the School of Computer and Software, Nanjing University of Information Science and Technology, China. He received the B.S. degree in Mathematical Science from Hunan Normal University and M.S. degree in Mathematical Science from Dalian University of Technology in 1984 and 1988, respectively. Then, he received the Ph.D. degree in Computer Engineering from Fudan University in 2001. His research interests include information security, network security, cryptography and ubiquitous computing security