Coverless Information Hiding Method Based on the Chinese Mathematical Expression

Xianyi Chen[1], Huiyu Sun[2], Yoshito Tobe[3], Zhili Zhou[1],
and Xingming Sun[1](✉)

[1] School of Computer and Software and Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology,
Nanjing 210044, Jiangsu, China
{0204622, zhou_zhili, sunnudt}@163.com

[2] Department of Computer Science, New York University, New York
NY 10012, USA
hs2879@nyu.edu

[3] Aoyama Gakuin University, Kanagawa 252-5258, Japan
yoshito-tobe@rcl-aoyama.jp

Abstract. Recently, many fruitful results have been presented in text information hiding such as text format-based, text image-based method and so on. However, existing information hiding approaches so far have been very difficult to resist the detecting techniques in steganalysis based on statistical analysis. Based on the Chinese mathematical expression, an efficient method of coverless text information hiding is presented, which is a brand-new method for information hiding. The proposed algorithm directly generates a stego-vector from the hidden information at first. Then based on text big data, a normal text that includes the stego-vector will be retrieved, which means that the secret messages can be send to the receiver without any modification for the stego-text. Therefore, this method is robust for any current steganalysis algorithm, and it has a great value in theory and practical significance.

Keywords: Coverless text information hiding · Chinese mathematical expression · Generation method · Steganography

1 Introduction

Information hiding is an ancient but also young and challenging subject. Utilizing the insensitivity of human sensory organs, as well as the redundancy of the digital signal itself, the secret information is hidden in a host signal, which does not affect the effect on sensory and value in use of the host signal. The host here covers all kinds of digital carriers such as the image, text, video and audio [1]. Since the text is the most frequently used and extensive information carrier, its research has attracted many scholars' interest, and has obtained many results.

There are four main types of text information hiding technologies: the text format-based, text image-based, generating method-based and embedding method-based natural language information hiding.

© Springer International Publishing Switzerland 2015

Z. Huang et al. (Eds.): ICCCS 2015, LNCS 9483, pp. 133–143, 2015.
DOI: 10.1007/978-3-319-27051-7_12

Text format based information hiding method mainly achieves the hiding of secret information by changing the character spacing, inserting invisible characters (spaces, tabs, special spaces, etc.) and modifying the format of documents (PDF, HTML, Office). For example, the [2, 3] hided data via changing the characters such as the row spacing, word spacing, character height and character width. The [4–8] embedded data that utilized the programming language to modify certain properties of the Office document (including NoProofing attribute values, character color attributes, font size, font type, font underline, Range object, object's Kerning property, color properties, etc.). Based on the format information of disk volume, Blass et al. proposed a robust hidden volume encryption in [9, 10]. The hiding capacity of information hiding methods based on text format is large, but most of them can't resist the attack of re-composing and OCR. They can't resist the steganography detection based on statistical analysis ([11, 12]).

The main idea of **information hiding method based on text image** is to regard a text as a kind of binary image. Then it combines the features of binary images with texts to hide data. For example, [13] embedded information by utilizing the parity of the numbers of black and white pixels in the block, in [14, 15], information was embedded by modifying the proportion of black-white pixels in a block and the pixel values of the outer edge, respectively. The embedding of secret information is realized by the rotation of the strokes of the Chinese characters in [16]. In addition, based on hierarchical coding, Daraee et al. [17] presented an information hiding method. Satir et al. [18] designed a text information hiding algorithm based on the compression method, which improved the embedding capacity. The biggest problem of text image based information hiding method is that it can't resist re-composing and OCR attacks. After re-composing the characters of the hidden information into a non-formatted text, the hidden information hiding would completely disappear.

Generation method based natural language information hiding method utilizes the natural language processing (NLP) technologies to carry secret information by generating the similar natural text content. It can be divided into two types: the primary text generation mechanism and the advanced text generation mechanism. The former is based on probability statistics, which is coded by utilizing the random dictionary or the occurrence frequency of the letter combinations, and then the generated text meets the natural language of statistical characteristics. The latter is based on the linguistic approach. Based on linguistic rules, it carries the secret data by using the imitation natural language text without specific content [19]. In these methods, due to the lack of artificial intelligence for the automatic generation of arbitrary text, the generation text always contains the idiomatic mistakes or common sense errors, or sentences without complete meaning. Moreover, it may cause incoherent semantic context and poor text readability, which is easily to be recognized by human eyes [20, 21].

Embedding method based natural language information hiding method embeds the secret information by using different granularity of modification of the text [22, 23]. According to the scope of the modified text data, the embedding method can be divided into lexical level information hiding and sentence level information hiding. The former method hides the messages by means of substitution of similar characters [24], substitution of spelling mistakes [25], substitution of abbreviations/acronyms and words in complete form [26], etc. Based on the understanding of the sentence structure and its semantics, the latter method changes the sentence structure to hide information, and then utilizes the syntactic transformation and restatement technology in the same situation of its meaning and style [27–30]. Embedding method is the focus and hotspot of text information hiding in current research. However, this method needs the support of natural language processing technology, such as syntactic parsing, disambiguation, automatic generation, etc., so that the information embedded into the text meets rationality of words, collocation accuracy, syntactic structure, and the statistical characteristics of language [31]. Because of the limitation of the existing NLP technology, it is hard to realize the hiding algorithm. In addition, there are still some deviation and distortion in the statistic and linguistics [32].

From the above we can see that the text information hiding has made many research results, but there are still some problems such as weak ability in anti-statistical analysis, bad text rationality and so on. Furthermore, theoretically as long as the carrier is modified, the secret message will certainly be detected. As long as the secret information exists in the cover, it can hardly escape from steganalysis. Thus, the existing steganography technology is facing a huge security challenge, and its development has encountered a bottleneck.

The proposed method firstly carries on syntactic parsing about the information to be hidden and divides it into independent keywords, then uses the Chinese mathematical expression [33] to create a locating tags. After that, utilizing the cloud search services-multi-keyword ranked search [34, 35], a normal text containing the secret information can be retrieved, which achieves the direct transmission of the secret information. It doesn't require any other carriers and modifications, while it can resist all kinds of existing steganalysis methods. This research has an important positive significance for the development of information hiding technology.

2 Related Works

The Chinese character mathematics expression was proposed by Sun et al. in 2002 [33]. The basic idea is to express the Chinese characters as a mathematical expression so that the operands are components of Chinese characters and the operators are six spatial relations of components. Some definitions are given below.

Definition 1. A basic component is composed of several strokes, and it may be a Chinese character or a part of a Chinese character.

Definition 2. An operator is the location relation between the components. Let A,B be two components, A lr B, A ud B, A ld B, A lu B, A ru B and A we B represent that A and B have the spatial relation of left-right, up-down, left-down, left-upper, right-upper, and whole enclosed respectively. An intuitive explanation of the six operators is shown in Fig. 1.

Definition 3. The priority of the six operators defined in Definition 3 is as follows: (1). () is the highest; (2). we, lu, ld, ru are in the middle; (3). lr, ud are the lowest; the operating direction is from left to right.

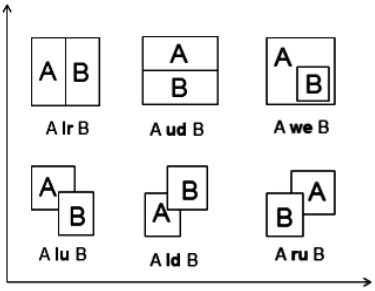

Fig. 1. Intuitive explanation of the defined operators

Using the selected 600 basic components and the six operators in Fig. 1, we can express all the 20902CJK Chinese characters in UNICODE 3.0 by utilizing the mathematical expressions. It is very nature and has a simple structure, and every character can be processed by certain operational rules as general mathematical expressions. After the expression of Chinese characters into the mathematical symbols, many processing of the Chinese information will become simpler than before.

According to the Chinese mathematical expression, we can see that if the appropriate components are selected as the location label of the secret message, it is better than that of the word or phrase being selected directly as the index in terms of many indicators such as randomness, distinguishability and universalness.

3 Proposed Method

Instead of conventional information hiding that needs to search an embedded carrier for the secret information, coverless information hiding requires no other carriers. It is driven by the secret information to generate an encryption vector, and then a normal text containing the encrypted vector can be retrieved from the big data of text, so the secret message can be embedded directly without any modification.

From the above analysis, there are three characteristics of the coverless information hiding algorithm: The first one is "no embedding", that is, a carrier can't embed secret information by modifying it. The second is "no additional message need to be transmitted except an original text", that is, other than the original agreement, there should not be any other carriers additionally used to send auxiliary information, such as the details or parameters of the embedding or extraction. The third is "anti-detection", which can resist all kinds of the existing detection algorithms. Based on the above characteristics together with the related theory of the Chinese mathematical expression, this paper presents a text-based coverless information hiding algorithm.

3.1 Information Hiding

For the coverless information hiding based on text, we first segment the secret data into words, then convert the Chinese words on a word-to-word basis, design the locating tags, generate the keywords that contain the converted secret data and the locating tags. Furthermore, we search the texts that contain the keywords in the database so as to achieve the information hiding with zero modification.

Let m be a Chinese character, and $\mathcal{T}$ be a set of the 20902CJK Chinese characters in UNICODE 3.0. Suppose the secret message is M = $\mathrm{m_1m_2...m_n}$, the conversion and location process of its secret information can be summarized as in Fig. 2. The details can be introduced as the following:

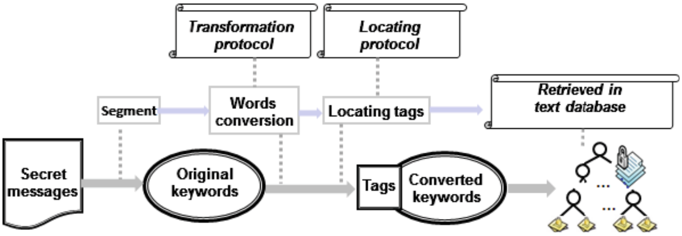

Fig. 2. Secret information conversion and positioning process

(1) *Information segmentation.* Based on the interdependence of the syntactic parsing, M is segmented into $\mathcal{N}$ non-overlapping keywords $w_i(i = 1, 2, \cdots \mathcal{N})$, where $||w_i|| \leq \ell$ and ℓ is a predetermined threshold for controlling the length of the keywords. The greater the ℓ, the higher the security, but the extraction of the secret information is more difficult. From the research of the Chinese information entropy, the choice of ℓ is often no more than three.

(2) *Words conversion.* Segment the text database retrieved into keywords according to the rules of information segmentation in step (1), and then calculate the frequency of every keywords in the text database, finally sort the words in descending order according to the frequency of occurrence $\mathcal{P} = \{p_1, p_2, \ldots, p_\mathrm{t}\}$. So the word transformation protocol can be designed as follows:

$$p_i' = \mathcal{F}_c(p_i, \mathrm{k}), i = 1, 2, \cdots, s;$$

Where $\mathcal{F}_c(p_i, \cdot)$ is the transformation function used for the statistic and analysis of the text database, this function is open for all users, and $\cdot$ is the private keys of the information receiver, such as the k of the above formula, and s is the number of keywords of the text database. Where the difference between the quantities of p_i and p_i' is not too much, if not, a commonly-used word will be converted into a rarely-used word, which will greatly decrease the retrieval efficiency of stego-text. Therefore, using the $\mathcal{F}_c$, the converted keyword $w\cdot_i$ of w_i can be calculated, we can obtain the converted secret message $\mathcal{W}\cdot = w\cdot_1 w\cdot_2 \ldots w\cdot_{\mathcal{N}}$.

(3) *Get the locating tags.* For the text database retrieved, divide the Chinese characters into various components by using the Chinese mathematical expression first, and then calculate the frequency of every component, finally sort the components in descending order according to the frequencies of occurrence. Select the component whose appearance time is in the top 50 and then determine the locating

sequence according to the user's key. For $i = 1, 2, \cdots \mathcal{N}$, suppose b_i is the corresponding component of the located keyword w'_i, when $\mathcal{N} > 50$, the keywords have the same located tags every 50 numbers.

For many components, the corresponding Chinese characters are often not unique. In order to find the stego-text that contains the secret information, we first calculate the r characters with the biggest numbers of appearance $m_i^j (j = 1, 2, \ldots, r)$ for every component b_i, combine every m_i^j with the keyword w'_i and research them from the text database, then sort m_j^i according to the number of occurrences. Utilizing the user's key, select the alternative character from the top 5 Chinese characters, so the location tags $\mathcal{L}_i (i = 1, 2, \cdots \mathcal{N})$ are calculated.

(4) Combining $\mathcal{L}_i$ with w'_i and obtain $\mathcal{D}_i (i = 1, 2, \cdots \mathcal{N})$, where $\mathcal{D}_i$ is the keyword retrieved in the text database.

In order to find the normal text that contains the keywords retrieved $\mathcal{D} = \{\mathcal{D}_i | i = 1, 2, \cdots \mathcal{N}\}$, the creation of large-scale text database plays a crucial role. It not only emphasizes of "high speed, wide range, great quantity", but also follows the principle of "quality, standardization and accuracy". Moreover, in order to improve the anti-detection performance, the quality of the text needs to be controlled from two aspects: One is to ensure that the text is normal with no secret information; the second is to ensure that the text is standardized in line with the language specification.

Based on the above text database, the keyword indexing technology is applied to find every $\mathcal{D}_i (i = 1, 2, \cdots \mathcal{N})$ in the database and built the reverse file index $\mathcal{ID}_i$, then search for a text that contains the secret information. If the search-string is found, then send it to the receiver; otherwise, divide it into two segments and re-retrieve it again until the right text is found. In order to avoid the suspicion, text classification can be used to the retrieval process, such as emotional and situational classification methods, which can avoid the retrieved results having non-relevant texts being grouped together.

It is worth mentioning that, from the above information hiding process we can see, the word conversion protocol is essentially a data encryption and the locating protocol is essentially a mapping. The two together realize the purpose of enhancing security and determining the location of the secret information. This idea isn't presented from nothing, but has a profound historical heritage. When choosing poetries as a text database, the keywords do not convert and the first word of each sentence is selected as the location tags, which is the ancient acrostic poem.

This skill is used in the peasant uprising famous novel "outlaws of the marsh" chapter sixty, which is about Chinese Northern Song Dynasty (1119-1121), such as Fig. 3. The normal reading order is from left to right, and its meaning is praising the general Lu Junyi. However, combine the initial letter in each line, we will get 卢俊义反, whose meaning is that "卢 will defect to the enemy"(卢 and 芦 is a homonym), thus achieving the purpose of hiding information in public document.

3.2 Information Extraction

In the conventional text information hiding, the stego-text is normal for a stranger but is abnormal for the receiver, so the receiver can extract the secret information by analyzing

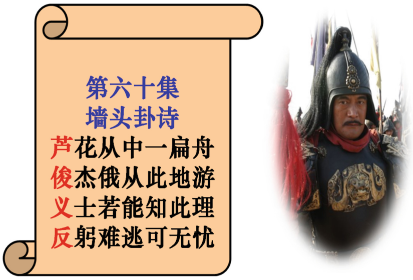

Fig. 3. Left is the acrostic and right is the picture of Lu Junyi.

the abnormalities. However, in the coverless information hiding, the stego-text is actually an open and normal text, and the receiver can't extract the secret information by finding the abnormal place. Let the stego-text be $\mathbb{S}$, and k is the private key, then the process of extraction is showed as Fig. 4. The details can be introduced as follows.

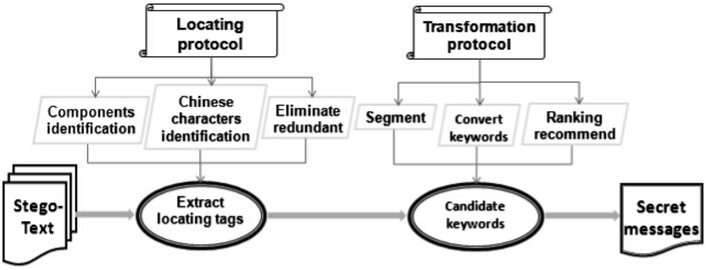

Fig. 4. The flowchart of the information extraction

(1) *Extraction preprocessing*. Because the text database is open to all users, they are also available as public information of the 50 Chinese components for marking and the corresponding Chinese characters, denoted $b = \{b_i | i = 1, \ldots, 50\}$ and $m_i = \{m_i^j | j = 1, \ldots\}$ respectively. Therefore, utilizing the user's private key, we can get the located component and its order of appearance. Moreover, based on the statistical results, it is also easy to get the Chinese characters $\mathcal{L} = \{\mathcal{L}_i | i = 1, 2, \cdots, \mathcal{N}\}$ in the database.

(2) *The extraction of candidate keywords.* Sequentially scan the stego-text $\mathbb{S}$, then extract the candidate locating tags $\mathcal{CL} = \{\mathcal{CL}_i | i = 1, 2, \cdots \mathcal{N}^{'}\}$ and the candidate keywords $\mathcal{S}' = \{\mathcal{S}_i' | i = 1, 2, \cdots \mathcal{N}^{'}\}$ according to the user location components set, where $\mathcal{N}^{'} \geq \mathcal{N}$.

When $\mathcal{N}^{'} = \mathcal{N}$, then $\mathcal{CL}$ is the locating tags of the secret information, skip step (3) to step (4); When $\mathcal{N}^{'} > N$, there exist the non-locating components that are contained in $\mathcal{CL}$, they should be eliminated from $\mathcal{CL}$ with step (3).

(3) *Eliminate the redundant tags*. The procedure is introduced as follows:

(a) Compare $\mathcal{CL}_i$ with $\mathcal{L}_i$ from $i = 1$, if $\mathcal{CL}_i = \mathcal{L}_i$, update $i = i + 1$ and execute the step (a) again, otherwise skip to step (b);

(b) If $\mathcal{CL}_{i-1} \neq \mathcal{CL}_i$, but both of them have the same components, then $\mathcal{CL}_i$ is not a location tag, delete it and skip to step (a). If $\mathcal{CL}_{i-1} \neq \mathcal{CL}_i$ and they have the different components, then compare the quantity sorting of the two Chinese characters in text database, and the Chinese character that doesn't meet the keys of receiver isn't the locating tag, then delete it from $\mathcal{CL}$; otherwise skip to step c);

(c) If $\mathcal{CL}_{i-1} = \mathcal{CL}_i$, then at least one character isn't the locating tag between $\mathcal{CL}_{i-1}$ and $\mathcal{CL}_i$, so combine $\mathcal{CL}_{i-1}$ and $\mathcal{CL}_i$ with its subsequent keywords to generate the keyword retrieved $\mathcal{D}_i$, delete the one with smaller number and skip to the step (a);

When the correct tags are calculated, locate it in $\mathbb{S}$, and then extract the character strings $\mathcal{S}_i (i = 1, 2, \cdots, \mathcal{N})$ after the locating points, where $||\mathcal{S}_i| = \ell$;

(4) Since each keyword is divided by the dependency syntax before the hiding, the length of every keyword $\ell_{\mathrm{w}i}$ may not be exactly the same, where $1 \leq \ell_{\mathrm{w}i} \leq \ell$. Moreover, because of the words conversion, the keywords cannot be accurately extracted. Therefore, when using the inverse transform of word conversion $\mathcal{F}_c^{-1}(p_i, k)$ to restore the string $\mathcal{S}_i$, the obtained candidate keywords set $\mathbb{K}_i = \left\{ \mathbb{K}_i^j | 1 \leq \mathrm{j} \leq \ell \right\}$ is not unique;

(5) Select a keyword from every $\mathbb{K}_i (i = 1, 2, \cdots \mathcal{N})$, and generate the candidate secret messages by researching the language feature and the word segmentation based on user background, then measure the confidence of the candidate secret information by analyzing the edit distance and similarity of the keywords, a rank can then be recommended to the receiver;

(6) Utilize the sorted recommended information, then combine the language analysis with Chinese grammar features, we can access the secret information $\mathrm{M} = m_1 \mathrm{m}_2 \ldots \mathrm{m}_\mathrm{n}$.

4 Example Verification

In order to clearly describe the above information hiding process, we explain it by a simple example. For example, let the secret information M be 无载体信息隐藏好, then the procedure of the hiding is shown in Fig. 5.

Firstly, segment M into $\mathscr{w}_1$ = 无, $\mathscr{w}_2$ = 载体, $\mathscr{w}_3$ = 信息, $\mathscr{w}_4$ = 隐藏, $\mathscr{w}_5$ = 好, then design the words conversion protocol $\mathcal{F}_c(p_i, \mathrm{k})$, where $\mathcal{F}_c(p_i, \mathrm{k})$ can be set to: "无→息,载体→一门,信息→古老, 隐藏→充满,好→的"..

Secondly, analyze the text database and calculate the statistic values, then choose the suitable component for the locating, where we set the components of the Chinese characters as "$\mathscr{b}_1$ =亻, $\mathscr{b}_2$ = 日, $\mathscr{b}_3$ =旡, $\mathscr{b}_4$ =勺, $\mathscr{b}_5$ =戈".

Thirdly, select the locating tags from the candidate Chinese characters, where we obtain the locating tags $\mathcal{L}$: "$\mathcal{L}_1$ =信、$\mathcal{L}_2$ =是、$\mathcal{L}_3$ =既、$\mathcal{L}_4$ =的、$\mathcal{L}_5$ =战", and the keywords set retrieved is $\mathcal{D}$ = {"信息", "是一门", "既古老", "的充满", "战的"}.

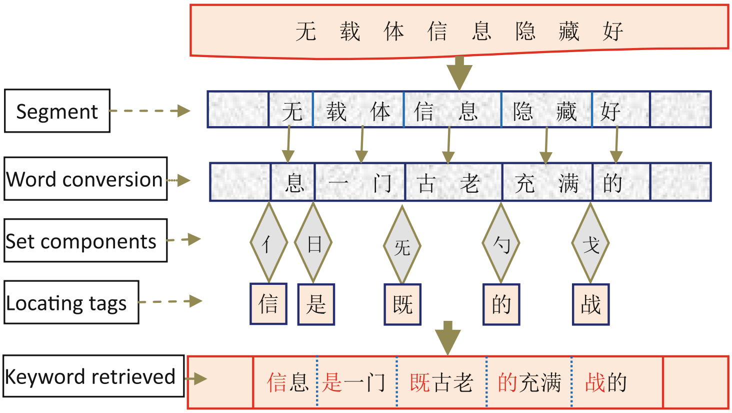

Fig. 5. Example results of the information hidding procedure

Finally, retrieve the text database to find a stego-text which contains the locating tags and keywords, where using the rules of the above, 信息隐藏是一门既古老又年轻的充满挑战的学科 is a stego-text with the retrieved secret information 无载体信息隐藏好.

In the case of the recipient's encrypted text information 信息隐藏是一门既古老又年轻的充满挑战的学科, because of the absence of redundant components, the extraction process is the inverse process of the embedding process. The realization is relatively simple, so we will not repeat them now.

5 Conclusions

This paper presented a text information hiding method, which is based on Chinese mathematical expression. Instead of the conventional information hiding method that needs to find an embedding carrier for the secret message, the proposed method requires no other carriers. First, an encryption vector is generated by the secret information, and then a normal text containing the encrypted vector is retrieved from the text database, which realizes embedding directly without any modification of the secret data. Therefore, the proposed method can resist all kinds of existing steganalysis methods. This research has an important positive significance for the development of information hiding technology.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (NSFC) (61232016, U1405254, 61502242, 61173141, 61173142, 61173136, 61373133), Jiangsu Basic Research Programs-Natural Science Foundation (SBK2015041480), Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (S8113084001), Open Fund of Demonstration Base of Internet Application Innovative Open Platform of Department of Education (KJRP1402), Priority Academic Program Development of Jiangsu Higher Education Institutions (PADA) Fund, Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) Fund, National Ministry of Science and Technology Special Project Research GYHY201301030, 2013DFG12860, BC2013012.

References

1. Cox, I.J., Miller, M.L.: The first 50 years of electronic watermarking. J. Appl. Signal Process. **2**, 126–132 (2002)
2. Low, S.H., Maxemchuk, N.F., Lapone, A.M.: Document identification for copyright protection using centroid detection. IEEE Trans. Commun. **46**(3), 372–383 (1998)
3. Brassil, J.T., Low, S.H., Maxemchuk, N.F.: Copyright protection for the electronic distribution of text documents. Proc. IEEE **87**(7), 1181–1196 (1999)
4. Ffencode for DOS (2015). http://www.burks.de/stegano/ffencode.html
5. WbStego4.2 (2015). http://home.tele2.at/wbailer/wbstego/
6. Kwan M. Snow (2015). http://www.darkside.com.au/snow/index.html
7. Koluguri, A., Gouse, S., Reddy, P.B.: Text steganography methods and its tools. Int. J. Adv. Sci. Tech. Res. **2**(4), 888–902 (2014)
8. Qi, X., Qi, J.: A desynchronization resilient watermarking scheme. In: Shi, Y.Q. (ed.) Transactions on Data Hiding and Multimedia Security IV. LNCS, vol. 5510, pp. 29–48. Springer, Heidelberg (2009)
9. Blass, E.O., Mayberry, T., Noubir, G., Onarlioglu, K.: Toward robust hidden volumes using write-only oblivious RAM. In: Proceedings of the 2014 ACM Conference on Computer and Communications Security (CCS 2014), pp. 203–214 (2014)
10. Mayberry, T., Blass, E.O., Chan, A.H.: Efficient Private file retrieval by combining ORAM and PIR. In: Proceedings of 20th Annual Network & Distributed System Security Symposium (NDSS 2014), pp. 1–11 (2014)
11. Goyal, L., Raman, M., Diwan, P.: A robust method for integrity protection of digital data in text document watermarking. Int. J. Sci. Res. Dev. **1**(6), 14–18 (2014)
12. Kwon, H., Kim, Y., Lee, S.: A tool for the detection of hidden data in microsoft compound document file format. In: 2008 International Conference on Information Science and Security, pp. 141–146 (2008)
13. Wu, M., Liu, B.: Data hiding in binary images for authentication and annotation. IEEE Trans. Multimedia **6**(4), 528–538 (2004)
14. Zhao, J., Koch, E.: Embedding robust labels into images for copyright protection. In: Proceedings of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Australia, pp. 242–251 (1995)
15. Xia, Z.H., Wang, S.H., Sun, X.M., Wang, J.: Print-scan resilient watermarking for the Chinese text image. Int. J. Grid Distrib. Comput. **6**(6), 51–62 (2013)
16. Tan, L.N., Sun, X.M., Sun, G.: Print-scan resilient text image watermarking based on stroke direction modulation for Chinese document authentication. Radioengineering **21**(1), 170–181 (2012)
17. Daraee, F., Mozaffari, S.: Watermarking in binary document images using fractal codes. Pattern Recogn. Lett. **35**, 120–129 (2014)
18. Satir, E., Isik, H.: A compression-based text steganography method. J. Syst. Softw. **85**(10), 2385–2394 (2012)
19. Wayner, P.: Disappearing Cryptography: Information Hiding: Steganography & Watermarking, 2nd edn. Morgan Kaufmann, San Francisco (2009)
20. Taskiran, C.M., Topkara, U., Topkara, M.: Attacks on lexical natural language steganography systems. In: Proceedings of the SPIE, Security, Steganography and Watermarking of Multimedia Contents VIII, San Jose, USA, pp. 97–105 (2006)
21. Meng, P., Huang, L.S, Yang, W.: Attacks on translation based steganography. In: 2009 IEEE Youth Conference on Information, Computing and Telecommunication, Beijing, China, pp. 227–230 (2009)

22. Nematollahi, M.A., Al-Haddad, S.A.R.: An overview of digital speech watermarking. Int. J. Speech Technol. **16**(4), 471–488 (2013)
23. Mali, M.L., Patil, N.N., Patil, J.B.: Implementation of text watermarking technique using natural language. In: IEEE International Conference on Communication Systems and Network Technologies, pp. 482–486 (2013)
24. Xiangrong, X., Xingming, S.: Design and implementation of content-based English text watermarking algorithm. Comput. Eng. **31**(22), 29–31 (2005)
25. Topkara, M., Topkara, U., Atallah, M.J.: Information hiding through errors: a confusing approach. In: Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents, San Jose, 6505 V (2007)
26. Rafat, K.F.: Enhanced text steganography in SMS. In: 2009 2nd International Conference on Computer, Control and Communication, Karachi, pp. 1–6 (2009)
27. Meral, H.M., Sankur, B., Ozsoy, A.S.: Natural language watermarking via morphosyntactic alterations. Comput. Speech Lang. **23**(1), 107–125 (2009)
28. Liu, Y., Sun, X., Wu, Y.: A natural language watermarking based on chinese syntax. In: Wang, L., Chen, K., Ong, Y.S. (eds.) ICNC 2005. LNCS, vol. 3612, pp. 958–961. Springer, Heidelberg (2005)
29. Kim, M.Y., Zaiane, O.R., Goebel, R.: Natural language watermarking based on syntactic displacement and morphological division. In: 2010 IEEE 34th Annual Computer Software and Applications Conference Workshops, Seoul, Korea, pp. 164–169 (2010)
30. Dai, Z.X., Hong, F.: Watermarking text documents based on entropy of part of speech string. J. Inf. Comput. Sci. **4**(1), 21–25 (2007)
31. Gang, L., Xingming, S., Lingyun, X., Yuling, L., Can, G.: Steganalysis on synonym substitution steganography. J. Comput. Res. Dev. **45**(10), 1696–1703 (2008)
32. Peng, M., Liu-sheng, H., Zhi-li, C., Wei, Y., Ming, Y.: Analysis and detection of translation based steganography. ACTA Electronica Sinica **38**(8), 1748–1752 (2010)
33. Sun, X.M., Chen, H.W., Yang, L.H., Tang, Y.Y.: Mathematical representation of a chinese character and its applications. Int. J. Pattern Recogn. Artif. Intell. **16**(8), 735–747 (2002)
34. Xia, Z., Wang, X., Sun, X., Wang, Q.: A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. IEEE Trans. Parallel Distrib. Syst. **99** (2015). doi: 10.1109/TPDS.2015.2401003
35. Fu, Z., Sun, X., Liu, Q., Zhou, L., Shu, J.: Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing. IEICE Trans. Commun. **98**, 190–200 (2015)