A Two-domain Data Hiding Scheme for Data Signature in Integrated Online Services Applications

Wei Yang¹, Caijun Zhang¹, Jinqiu Zhang², Kaiqiang Xian¹, Qianjun Wu³, Xiaolong Wang³*

¹State Grid Customer Service Center, Hexi District, Tianjin 300200, China
²Marketing Service Center, State Grid Jiangsu Electric Power Co., LTD., Nanjing 211000, China
³State Grid Electric Power Research Institute, Nanjing 211000, China.
*Corresponding author: wangxiaolong2@sgepri.sgcc.com.cn

Abstract. There is confidential data in integrated online services applications. The data needs to be hidden before transmission on the Internet, and the data authenticity should be assured. Information hiding technology is an effective way of protecting the confidential data and may be used for assuring data authenticity. A Two-domain data hiding scheme (TDHS) is introduced in the paper to protect contract document authenticity in grid system. TDHS aims to embed confidential data into signature image in both spatial and frequency domain. LSB-like algorithm is used to generate a signature image $H^{\text{sp}}$ by embedding feature data of contract document into the original signature image $H$ in the spatial domain. Then DCT algorithm is adopted to embed the watermark in signature image $H^{\text{sp}}$ in the frequency domain, and generate signature image $H^{\text{freq}}$. The signature image $H^{\text{freq}}$ is inserted into the contract document. When received the document with signature, the recipient can verify its authenticity. The feasibility of TDHS is verified through experiments.

1. Introduction

There are many contract documents in the power grid business systems. These documents need to be signed remotely and may have some confidential data. The contract documents signing is a process of mutual trust between both parties. However, there are some people who maliciously use other people's signatures to forge the contents of the contract documents for profit. Therefore, an effective method is necessary for providing authenticity of the contract document signature and preventing signatures from being fraudulently used.

In this case, the information hiding/data hiding/steganography technology can provide a good protection method for the electronic signature image. Information hiding is a technology that embeds data in digital media for copyright protection or secret communication [1]. Compared with encryption technology, multi-form data can be unknowingly transmitted to the recipient [2]. The cover of information hiding is multimedia data, including images, text, audio, video, etc. Since images are one of the most important media in power grid business systems, a lot of emphasis has been placed on image steganography [3]. Digital watermarking technology is to embed specific digital information (identity information, serial numbers, text or image logos, etc.) into media data to achieve privacy protection, information security or copyright protection.
Traditional image steganography can be divided into two categories: methods based on spatial domain and methods based on transform domain [4]. For the method based on the spatial domain, the secret information is directly embedded in the image pixel value. Typical methods include least significant bit (LSB) [5], prediction error [6], histogram-based method [7], modular operation [8], quantization-based method [9] and other methods. Algorithms based on the spatial domain have little effect on the quality of the cover image, and the amount of embedding is large. However, it usually has poor robustness. In order to improve the robustness, researchers proposed to embed secret information into the transform domain. Typical methods include steganography based on Discrete Cosine Transform (DCT) [10], Discrete Wavelet Transform (DWT) [11] and Discrete Fourier Transform (DFT) [12].

This paper presents a two-domain data hiding scheme (TDHS) for data signature in grid system, which combines hand signature, feature data embedding in spatial domain and digital watermark embedding in frequency domain.

The innovations of this paper are as follows:
1) A Two-domain Data Hiding scheme is designed for hiding confidential data in contact document signature in power grid business system and for verifying the authenticity of the signature;
2) In TDHS, the feature data of the contract text and a watermark is embedded in spatial and frequency domain respectively, in order to hide data and verify authenticity effectively;
3) The feasibility of TDHS is verified by an experimental implementation in detail.

The paper is organized as follows. The structure of TDHS is introduced in Section 2. The document feature data and embedding algorithm in spatial domain and watermark embedding algorithm in frequency domain are presented in Section 3 and Section 4 respectively. An experimental implementation is described in Section 5. Finally, Section 6 concludes the paper.

2. Two-domain Data Hiding scheme (TDHS)
TDHS is designed for provide authenticity of the contract document signature and prevent signatures from being fraudulently used. The structure and procedure of TDHS is shown in Figure 1.

The signatory extracts the text feature value of the contract document. The feature data may be the frequency or position of the text in the text message [14]. The feature data is embedded into the signature image \( H_0 \) from a signatory, in the spatial domain, and a new image \( H_0 \) is obtained. Since the feature value of the document only corresponds to the document to which it belongs, this information can be used to check whether the content has been modified.

Next, the image \( H_0 \) is transformed into a frequency domain image using discrete cosine transform (DCT). The watermark image is also transformed into a frequency domain image with DCT. Then the two images are merged into a new image imageA in frequency domain. The imageA is transformed into a spatial domain image (\( H_{ready}^s \)) through inverse discrete cosine transform [13].

Finally, image \( H_{ready}^s \) is inserted into contract document and the signing process is finished.

When received the signed contract document, the recipient firstly separates the contract document file and the signature image \( H_{ready}^s \). Then, the document feature data A of the electronic contract document is calculated from the contract document file. Next, the feature data B is recovered from \( H_{ready}^s \). The two kinds of feature data, A and B, are compared to verify whether they have the same value.

After that, the signature image \( H_{ready}^s \) is transformed into a frequency domain image by discrete cosine transform, the watermark image is extracted in frequency domain. Then the recovered watermark image is compared with the stub watermark image to verify whether it is acceptable.

If the above two-stage verification is passed, the contract document can be accepted.
3. **Document Feature Value and Spatial Embedding Algorithm**

The document feature data is the feature value calculated from the contract document [15]. Just like the human gene, the document feature data may be also unique. For different documents, the frequency of use of the text, the order of the text, and the grammatical structure will be different. The difference can be regarded as the feature data for recognition of the document. The selection of the feature data can be the frequency of one or more characters in the entire document, or one or more characters in the entire document.

Spatial embedding algorithm is an embedding algorithm that can embed large-capacity data into an image. It has the characteristics of simple embedding process, but has low robustness. Once it is maliciously attacked, the embedded information will be difficult to recover. The commonly used embedding algorithm in spatial domain is the least significant bit (LSB) algorithm which is adopted in this paper.

Embedding the feature data of the document into the image using the spatial embedding algorithm is like to assigning a unique identification to the document [16]. Whenever the content is modified, inconsistencies will be detected in the verification process.

Steps for embedding document feature values in spatial domain is as follows.

1) The signatory prepares the contract document to be signed, and provides the corresponding original signature image \( H \), and uses the extraction algorithm to extract the feature data of the document;

2) The width and height of the original electronic signature image \( H \) are adjusted to an integer multiple of \( M \), and the adjusted image is divided into blocks according to \( M \times M \) to obtain a total of \( n \) blocks. In this paper, \( M = 8 \);

3) The feature data is binarized and converted into a binary string, \( n \) bits are selected from the string, and the spatial embedding method is adopted to embed each image block separately;

4) Combine the above \( n \) images with embedded information together to form a signature image \( H^S \) after the feature data is embedded.

The process of embedding document feature values in spatial domain is shown in Figure 2.
Figure 2. Process of embedding in spatial domain

4. Frequency Domain Watermark Embedding Algorithm

When the signature image $H^s$ is ready, the watermark image is to be embedded in the frequency domain. The watermark image $w$ is an image that is unique [17].

The frequency domain watermark embedding adopts the discrete cosine transforms method. The DCT formula is as follows.

$$ F(m,n) = c(m)c(n) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{\pi(2x+1)m}{2M} \cos \frac{\pi(2y+1)n}{2N} $$

(1)

Where,

$$ m = 0,1,\ldots; n = 0,1,\ldots,N-1 $$

$$ c(m) = \begin{cases} 
\frac{1}{M} & m = 0 \\
\frac{2}{M} & m = 1,2\cdots M-1
\end{cases} $$

$$ c(n) = \begin{cases} 
\frac{1}{N} & m = 0 \\
\frac{2}{N} & n = 1,2\cdots N-1
\end{cases} $$

The inverse discrete cosine transform formula is:

$$ f(x,y) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} c(m)c(n)F(m,n) \cos \frac{\pi(2x+1)m}{2M} \cos \frac{\pi(2y+1)n}{2N} $$

(2)

Where, $x,y$ are spatial sampling values, $m,n$ are frequency domain sampling values and $m = 0,1,\ldots; n = 0,1,\ldots,N-1$.

If $M = N$, the positive and negative transformations of the discrete cosine transform can be simplified as:

$$ F(m,n) = c(m)c(n) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{\pi(2x+1)m}{2M} \cos \frac{\pi(2y+1)n}{2N} $$

(3)

$$ f(x,y) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} c(m)c(n)F(m,n) \cos \frac{\pi(2x+1)m}{2M} \cos \frac{\pi(2y+1)n}{2N} $$

(4)

Where,
\[ m = 0,1,\cdots ; n = 0,1,\cdots , N - 1 \]
\[ c(m) = c(n) = \begin{cases} 
1 & m = 0 \text{ or } n = 0 \\
\frac{1}{\sqrt{2}} & m, n = 1,2,\cdots , N - 1 
\end{cases} \]

Read the signature image \( H^s \) embedded in the airspace

Get \( H_m \) after processing

Use DCT to convert \( H^s_m \) to \( H^s_{m-DCT} \)

Get \( w_n \) after processing

Insert \( w_n \) into the coefficient of \( H^s_{m-OCT} \)

\( H^s_{ready} \)

**Figure 3. Process of embedding in frequency domain.**

Steps of the algorithm is as follows.

Let \( H^s_{m} \) be the image subdivided from \( H^s \) and \( w, H^s_{m-DCT} \) is the image converted from \( H^s_{m} \) after DCT transformation, and \( H^s_{m-F} \) is the image in the frequency domain \( H^s_{m-DCT} \) and \( w_n \) combined image.

1) Read the signature image \( H^s \) and the watermark image \( w \) after the airspace information is embedded;

2) The signature image \( H^s \) is divided into a combination of \( 8 \times 8 \) image blocks that do not cover each other;
\[ H^s = \{ h^s(i,j), 0 \leq i, j < M \} \]
\[ H^s_m = \{ h^s_m(i,j), 0 \leq i, j < 8 \} \]

3) Divide the watermark image \( w \) into a combination of \( 2 \times 2 \) blocks.
\[ w = \{ w(i,j), 0 \leq i, j \leq M \} \]
\[ w_n = \{ w_n(i,j), 0 \leq i, j < 2 \} \]

where \( w_n(i,j) \in \{0,1\} \), \( n \) is the total amount of \( 2 \times 2 \) blocks;

4) Convert \( H^s_m \) to \( H^s_{m-DCT} \) by DCT;

5) Insert \( w_n \) into the coefficient of \( H^s_{m-DCT} \).
\[ H^s_{m-F} = \{ h^s_{m-F}(i,j) = h^s_{m-DCT}(i,j) \oplus w_n(i,j), 0 \leq i, j < 8 \} \]

6) Convert the embedded host image \( H^s_{m-F} \) by the inverse transformation of DCT;

The process of embedding a watermark image in the frequency domain is shown in Figure 3.

**5. Experimental results**

The experimental implementation is carried out according to the embedded algorithms described above. Python is used to write the simulating program. The hardware and software environment used in the experiment is CPU: Intel(R) Core (TM) i7-10700 @ 2.90 GHz; RAM: 16 GB; OS: Windows 10; Programming: Python and PyCharm 2020.2.3.
Product Cooperation Agreement

In accordance with relevant laws, administrative regulations and relevant management regulations, based on the principles of equality, mutual benefit, and voluntary cooperation, after consensus, the two parties agree to reach the following agreement in accordance with the negotiation provisions of multi-party negotiations, and the parties shall abide by it.

Signatory:

Figure 4. Image of the document to be signed

The contract document to be signed and transmitted is shown in Figure 4. A watermark image $w$ in this experiment is shown in Figure 5.

Three images with different sizes, different backgrounds are used as signature image $H$ in the experiment, as shown in Figure 6.

![Signature images](image)

Figure 6. Signature images

The feature data of contract document is embedded in signature image $H$ in spatial domain, and then a new image $H^s$ is obtained.

The watermark image $w$ is embedded in image $H^s$ in frequency domains, and then a new signature image $H^s_{\text{ready}}$ is generated. Image $H^s_{\text{ready}}$ is inserted into the contract document, as shown in Figure 7. At this time, the data hiding and signing process is finished. The contract document obtained is ready for transmitting from the signatory to recipient.

![Electronic signature image with double hidden information](image)

Figure 7. Electronic signature image with double hidden information

Upon signed contract document is received, the recipient firstly separates the contract document file and the signature image $H^s_{\text{ready}}$. Then, the document feature data $A$ of the electronic contract document
is calculated from the contract document file. Next, the feature data B is recovered from $H_{ready}^S$. The two kinks of feature data, A and B, are compared to verify whether they have the same value, as shown in Figure 8 a).

![Figure 8. The extracted watermark image](image)

The signature image $H_{ready}^S$ is transformed into a frequency domain image by discrete cosine transform, the watermark image is extracted in frequency domain, as shown in Figure 8 b). Then the recovered watermark image is compared with the stub watermark image to verify whether it is acceptable.

The watermark and feature value comparisons give an acceptable result, so the signed contract document is accepted.

### 6. Conclusion

Multi-form data exists in the power grid business system. Some data may like contract documents contain critical information for business. The contract documents need to be signed and require verification of authenticity. This paper present TDHS, an effective method, to provide authenticity of the contract document signature and prevent signatures from being fraudulently used.

TDHS embeds contract document feature data into signature image in spatial domain. LSB-like algorithm is used to complete the data hiding process and generate a signature image $H^S$. Then DCT algorithm is used to embed the watermark in signature image $H^S$ in the frequency domain, and generate signature image $H_{ready}^S$. The signing work is done by inserting signature image $H_{ready}^S$ into the contract document. TDHS makes the recipient easy to verify the authenticity of contract documents with signature.

We implanted a prototype TDHS system, which shows the feasibility of our design. In the future, we will make further improvements to TDHS and enhance its performance.

### Acknowledgement

This research was financially supported by the Science and Technology projects of State Grid Corporation of China (Contract No. 5700-202153179A-0-0-00 / Research on Digital Mutual Trust and Cross-domain Sharing Technology for Integrated Online Electric Power Services Applications).

### References

[1] N. F. Johnson and G. Mason, “Exploring Steganography: Seeing the Unseen,” 1998.

[2] C. X. Shen, H. G. Zhang, D. G. Feng, Z. F. Cao, and J. W. Huang, “Survey of information security,” Sci. China, Ser. F Inf. Sci., vol. 50, no. 3, pp. 273–298, 2007, doi: 10.1007/s11432-007-0037-2.

[3] S. C.P, S. T, and U. G, “A Study of Various Steganographic Techniques Used for Information Hiding,” Int. J. Comput. Sci. Eng. Surv., vol. 4, no. 6, pp. 9–25, 2013, doi: 10.5121/ijcses.2013.4602.

[4] Y. Zhang, J. Jiang, Y. Zha, H. Zhang, and S. Zhao, “Research on Embedding Capacity and Efficiency of Information Hiding Based on Digital Images,” Int. J. Intell. Sci., vol. 03, no. 02, pp. 77–85, 2013, doi: 10.4236/ijis.2013.32009.
[5] H. B. Kekre, V. P. West, R. Khanna, and A. Hussaini, “Comparison between the basic LSB Replacement Technique and Increased Capacity of Information Hiding in LSB’s Method for Images,” Int. J. Comput. Appl., vol. 45, no. 1, pp. 33–38, 2012.

[6] A. M. Molinaro, R. Simon, and R. M. Pfeiffer, “Prediction error estimation: a comparison of resampling methods,” vol. 21, no. 15, pp. 3301–3307, 2005, doi: 10.1093/bioinformatics/bti499.

[7] W. Hong, “Adaptive reversible data hiding method based on error energy control and histogram shifting,” Opt. Commun., vol. 285, no. 2, pp. 101–108, Jan. 2012, doi: 10.1016/j.optcom.2011.09.005.

[8] F. A. Jassim, “A Novel Steganography Algorithm for Hiding Text in Image using Five Modulus Method,” Int. J. Comput. Appl., vol. 72, no. 17, pp. 975–8887, 2013.

[9] C. Chang, “Information Hiding in Dual Images with Reversibility,” 2009, doi: 10.1109/MUE.2009.35.

[10] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, “A DCT-domain system for robust image watermarking,” Signal Processing, vol. 66, no. 3, pp. 357–372, May 1998, doi: 10.1016/S0165-1684(98)00015-2.

[11] A. M. Eskicioglu, “Robust DWT-SVD Domain Image Watermarking: Embedding Data in All Frequencies,” pp. 166–174.

[12] V. Solachidis, I. Pitas, and S. Member, “Circularly Symmetric Watermark Embedding in 2-D DFT Domain,” vol. 10, no. 11, pp. 1741–1753, 2001.

[13] M. M. Yeung and F. Mintzer, “Invisible watermarking technique for image verification,” IEEE Int. Conf. Image Process., vol. 2, pp. 680–682, 1997, doi: 10.1109/icip.1997.638587.

[14] L. M. Abualigah and A. T. Khader, “Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering,” J. Supercomput., vol. 73, no. 11, pp. 4773–4795, Nov. 2017, doi: 10.1007/s11227-017-2046-2.

[15] I. Chalkidis, I. Androutsopoulos, and A. Michos, “Extracting contract elements,” Proc. Int. Conf. Artif. Intell. Law, pp. 19–28, 2017, doi: 10.1145/3086512.3086515.

[16] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, “Multimedia data-embedding and watermarking technologies,” Proc. IEEE, vol. 86, no. 6, pp. 1064–1087, 1998, doi: 10.1109/5.687830.

[17] C. T. Hsu and J. L. Wu, “Hidden digital watermarks in images,” IEEE Trans. Image Process., vol. 8, no. 1, pp. 58–68, 1999, doi: 10.1109/83.736686.