Secret Sharing with Steganographic Effects for HTML Documents

Kuei-Li Huang and *Wen-Hsiang Tsai*
Department of Computer and Information Science
National Chiao Tung University
Hsinchu, Taiwan 300, Republic of China
e-mail: gis91568@cis.nctu.edu.tw, whtsai@cis.nctu.edu.tw

ABSTRACT

A secret sharing method with steganographic effects for HTML documents is proposed in this study. Secret sharing is a method for information protection, which translates a secret into several shares for the participants to keep, and recovers the secret by collecting a pre-defined number of the shares. In the secret sharing process, important components in an input secret HTML document are extracted and shared by cooperative sharing operations with data magnitude control. In the process of steganographic effect creation for the shares, each component in each share is independently transformed into a fake version of the same component type. After the process, each share is still of the same style of the input secret HTML document, thus achieving an overall steganographic effect. In the secret recovery process, after collecting all the HTML-type shares, each component of the secret HTML document can be recovered using the shares extracted from the corresponding components of the HTML-type shares. Experimental results show the feasibility of the proposed method.

1. INTRODUCTION

Secret sharing is a way to encrypt and distribute secret information into several shares so that each share, kept by a participant, contains only partial information of the secret. The secret data can be recovered if a pre-defined group of shares is collected. Moreover, collecting a group of shares different from the pre-defined group cannot recover the secret information.

Shamir [1] was the first to propose the concept of secret sharing in his (k, n)-threshold method, where n denotes the number of participants and the threshold k specifies the minimum number of shares in the pre-defined group. By this method, secret information is encrypted and then distributed into n shares, which are then assigned to n participants, respectively. If and only if k or more than k participants get together, the secret information can be *recovered* by a certain method. Subsequently, many related topics were studied [2]-[7] and various kinds of secret sharing methods were proposed [8]-[13]. Nevertheless, these proposed methods are only suitable for data of short lengths, such as passwords, encryption and decryption keys, and so on.

Based on ideas from company organizations, Lin and Tsai [14] proposed a method of hierarchical secret sharing, which is a new concept of sharing secret information among groups of participants. Three types of secret sharing, cooperative sharing, independent sharing and dominant sharing, were proposed for realizing the concept. The method first specifies a hierarchical structure as a tree, in which each non-leaf node denotes one of the three operations and each leaf node denotes one of the shares. According to the tree, secret information can be encrypted and distributed into shares corresponding to the leaf nodes and recovered as well.

A steganographic method for text-type documents, which is also investigated in this study, means a certain way to embed message data into a text-type document to avoid awareness of the message. The capacity of the redundancy information of text-type documents for hiding data is smaller than those of images or videos. Methods for hiding data in text-type documents may embed data into the text itself or in the language describing the text format. For documents of complicated formats, Chang and Tsai [15] proposed a method for covert communication using HTML documents. Because a web browser does not display a sequence of spaces following a leading space and tags in an HTML document, the method can encode secret information by adjusting the sizes of between-word spaces and the expressions of the tags in the document.

In this study, we propose a secret sharing method for secret HTML documents with additional steganographic effects to create a practical application on secret text sharing. While doing secret sharing, only the important components in a secret HTML document are encoded and distributed into shares by cooperative sharing operations with a technique of data magnitude control proposed in this study. This data magnitude control technique bounds the data magnitude of a share so that the cooperative sharing operation can be applied appropriately. A steganographic method also proposed in this study then translates the shares of the components into types similar to those of the original components, which makes the shares meaningful. Therefore, a meaningful share, an HTML-type share, results which can prevent from illicit users' curiosity to

a certain degree, compared with a "raw" share, a share obtained without applying steganographic techniques.

The remainder of this paper is organized as follows. In Section 2, the proposed secret sharing method is presented. The proposed steganographic method is described in Section 3. In Section 4, the recovery process is described. And several experimental results are illustrated in Section 5. Finally, some concluding remarks as well as some suggestions for future works are stated in Section 6. In the appendix, the proof of an equation used in the secret sharing method is included.

2. PROPOSED SECRET SHARING METHOD

The proposed strategy for sharing HTML documents is to extract the important components in a secret HTML document and conduct secret sharing, which retains the component framework of the secret HTML document for steganography. For secret sharing, the cooperative sharing operation is adopted, accompanied with the proposed data magnitude control technique.

2.1. Cooperative Sharing with Data Magnitude Control

In the proposed technique of controlling the data magnitude of the shares, the concept of *modulus* is applied to the cooperative sharing operation originally proposed by Lin and Tsai [14], resulting in the following new operation:

$$f(x) = \left(\sum_{i=1}^{n} (x - a_i) + s \right)_{\bmod p}, \qquad (1)$$

where p is a prime number, n is the number of participants, s is a secret in the form of an integer, and all a_i are randomly selected integers, with $i = 1, 2, \ldots, n$. The ith participant takes the ith share $(a_i, f(a_i))$ which is a pair of integers. In this study, p is set to the prime number 257 so that each share resulting from (1) above becomes a pair of *8-bit integers or the exceptional values of* 256, if we always take s and all a_i in (1) to be an integer in the range of {0, 1, …, 255}, i. e., to be a byte. To deal with the exceptional case of 256 so that $f(x)$ can be bound to the range of 0 through 255, an additional constraint imposed on the selection of the values of a_i is designed in this study for use during the secret sharing and is described as follows:

$$a_1 + a_2 + \ldots + a_n = [(n - 1) \times 256 + s]_{\bmod 257}. \qquad (2)$$

The verification that the constraint limits $f(x)$ to the range of 0 through 255 is described in the Appendix. By controlling the data magnitude of shares in the above way, a share $(a_i, f(a_i))$ can be stored with a two-byte space while a secret s a one-byte space.

When all the shares are collected, s can be obtained by the following recovery function:

$$s = \left(f(a_i) - (n-1) \times a_i + \sum_{k=1; k \neq i}^{n} a_k \right)_{\bmod 257}. \qquad (3)$$

2.2. Secret Sharing Process

The secret sharing process first classifies the components in a secret HTML document into two types: text and non-text, and extracts the important parts of the components. Then, the shares of the important part of each component are generated by the above-mentioned revised cooperative sharing operation with modulus 257. Figure 1 shows a flowchart of the process.

3. PROPOSED STEGANOGRAPHIC METHOD

Two steganographic techniques for components of the two types, text and non-text, respectively, are proposed in this study, and are described subsequently.

3.1. Text Component Steganography

According to a property of the HTML document, a text component that can be seen on browsers can be substituted by a fake text component and one of the shares of the text component are so translated and hidden in the substitute one in this study without arousing any awareness of the share. In the following, the behavior of text contents on a browser is described. Then, the proposed steganographic technique is described.

The text components in an HTML document mean those texts outside tags and can be displayed on a browser. Only one space symbol between two successive words is displayed on a browser while actually a string of three types of symbols, namely, the tab, new-line, and space symbols, are bundled into the gap between two consecutive words in an HTML document. An additional ANSI code 0x0C has also been found in this study to possess the same function as the three types of symbols on the Internet Explore (IE) browser. For other browsers, there might exist other symbols or codes with the same property as the above-mentioned three types of symbols.

These symbols are used in the proposed steganographic technique. In this study, the IE browser is used as the HTML browser.

By utilizing the property mentioned above, we can now describe the proposed steganographic method for text component shares. A share of a text component is first translated into a string made up of the previously-mentioned four types of symbols and then equally segmented and embedded into the between-word spaces of a fake text component. The fake text component can be obtained from an article and

is of the length approximately identical to that of the original text component.

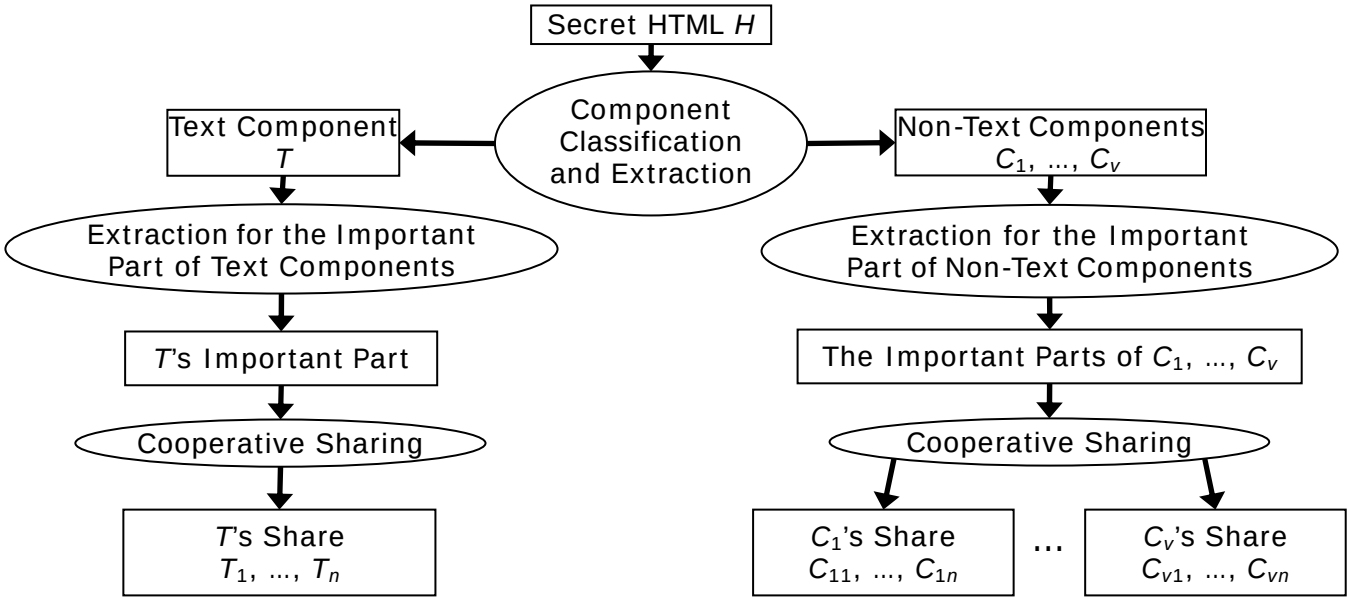


Figure 1: The flowchart of the secret sharing process.

3.2. Non-Text Component Steganography

Non-text components that can be displayed on browsers are of the form of links in the tags of an HTML document. According to the concept of "dynamic" links, in the proposed steganographic method for a non-text component, it is replaced with a fake link first and a share of the non-text component becomes part of the fake link. In the following, how a dynamic link works on the internet is introduced first, how to replace a link into a workable dynamic one that contains a share of the original component is proposed then, and, in the end, an illustrative example is given.

A "static" link is the link which indicates directly the address of the corresponding source on the internet. However, a "dynamic" link is a link which contains the address of an agent in a server on the Internet, followed by the information of the agent. A dynamic link, for example, is of the form: "http:// www.cis.nctu.edu.tw/ ~gis91568/agent?123456," where the string before the question mark is the address of the agent and the string following the question mark is the information for the agent. After a browser gets the address of the agent from the link and informs the agent the information in the link, the agent returns the corresponding source according to the information. For example, a dynamic image link containing the address of an image database agent and the information of the image index in the database can let a browser know where the image database agent is on the Internet and which image should be retrieved from the database according to the image index. Finally, the agent returns the image with the specified index in the image database through the Internet to the browser.

In our method, we first create some agents for links of different types (e.g., audio, video, flash, image, etc.) and put the agents onto a server. Then, according to the type of the non-text component in concern, a dynamic link is created by concatenating the address of the corresponding agent with a question mark followed by some information for the agent. The information put at the rear of the dynamic link is the string of a share of the component in the hexadecimal form.

After the above two steganographic processes, several HTML-type shares of the same style as that of the secret HTML are generated. That is, the component framework of an HTML-type share is the same as that of the secret HTML.

4. RECOVERY PROCESS

When a sufficient number of HTML-type shares are collected, the recovery process for the secret HTML document can be launched, as shown in Figure 2.

The first step of the process is to extract all the shares of each original HTML component in the components of all the HTML-type shares. Then, related shares are grouped and processed to obtain a component of the secret HTML document using (3). After this step, all the components of the secret HTML document are recovered. Finally, by referring to the component framework of an HTML-type share and replacing the fake components of the share with the just recovered components, the original HTML document can be recovered.

5. EXPERIMENTAL RESULTS

Suppose that the number of secret sharing participants is two. Figure 3 shows a secret HTML which contains two text component segments, “This is a secret sharing test for HTML.” at the top and at the bottom, a hyperlink with hyperlink text string “This is a secret sharing link.”, an image at the middle left, a video component at the middle center, and a Flash file at the middle right. The two pictures in Figure 4 are the resulting HTML-type shares. The component framework of each HTML-type share is the same as that of the secret HTML document, while the components in each HTML-type share are different from those in the secret HTML document. The HTML document in Figure 5 is the recovered secret HTML which is the same as the original secret HTML seen on browsers. Figure 6(a) shows an image source path, the secret image component of the secret HTML, and the two corresponding stego-image components are shown in Figures 6(b) and (c). In the image component, the string before the question mark is the address of the image agent and the under line string following the question mark is one of the translated share data of the secret image component.

6. CONCLUSION

Because the important parts of a secret HTML document are the components that can be displayed or be accessed on browsers, it is proposed in this study to share these components among participants of the secret HTML document by the cooperative sharing operation with data magnitude control by the modulus operation. In order to create steganographic effects on the shares of the components, two steganographic techniques for the text component and the non-text components of secret HTML documents are proposed. For a share of a text component, the proposed technique substitutes the original text component by an article with the share hidden into between-word spaces. For a share of a non-text component, the proposed technique uses a dynamic link with the share as the parameter of the link to create steganographic effects. After applying the two steganographic techniques to the shares of the components in the secret HTML document, HTML-type shares with styles identical to that of the secret HTML document are generated. Experimental results show the applicability of the proposed methods to real HTML documents.

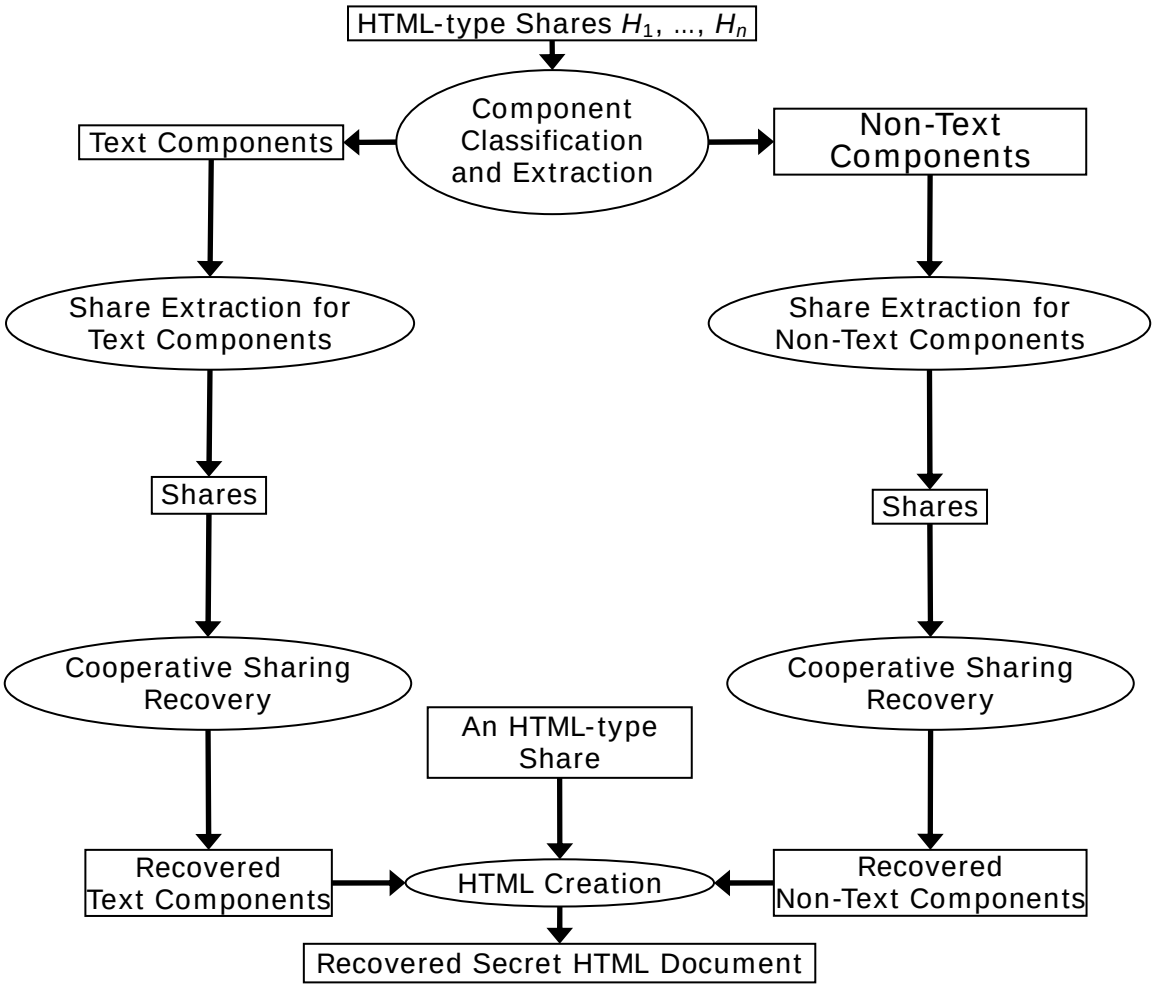


Figure 2: Flowchart of the recovery process.

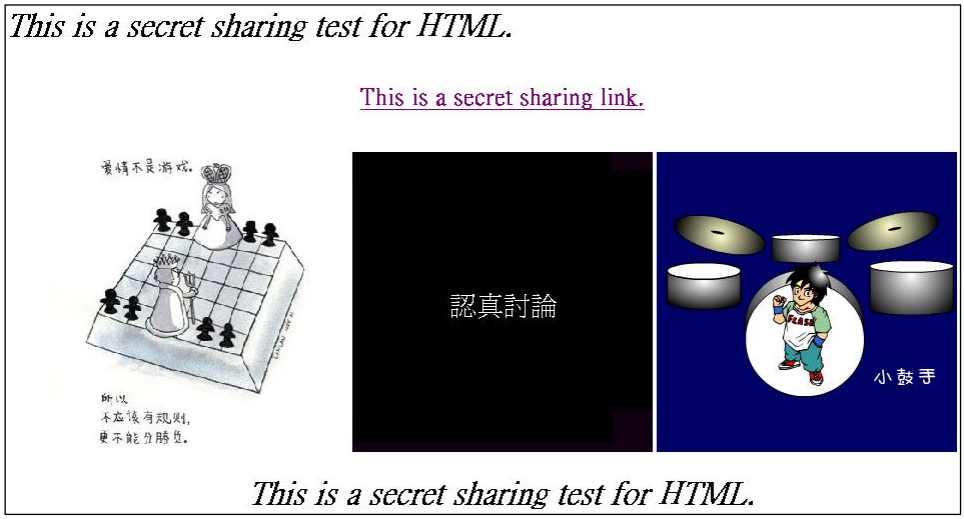
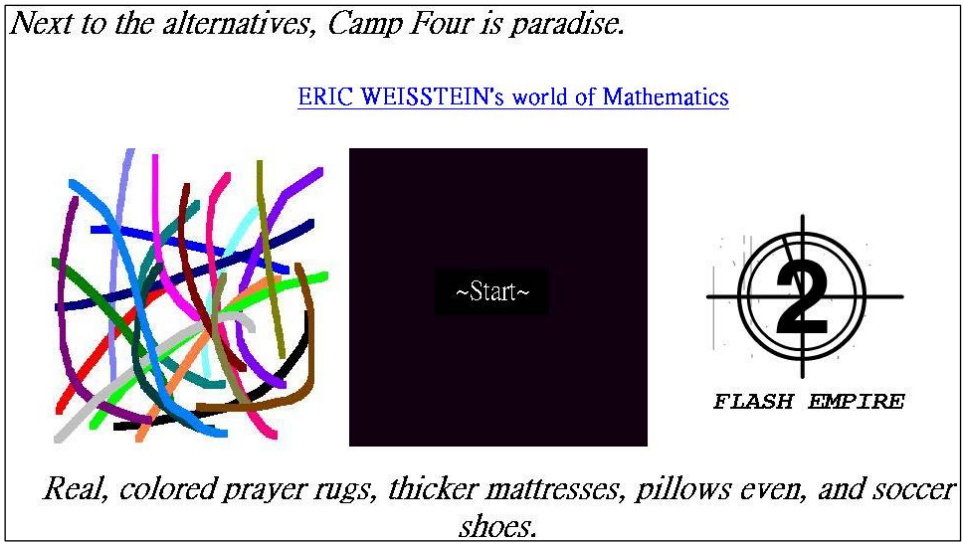


Figure 3: A secret HTML document.

(a)



(b)

Figure 4: HTML-type shares (a) through (b) HTML-type shares generated from the secret HTML document in Figure 3.

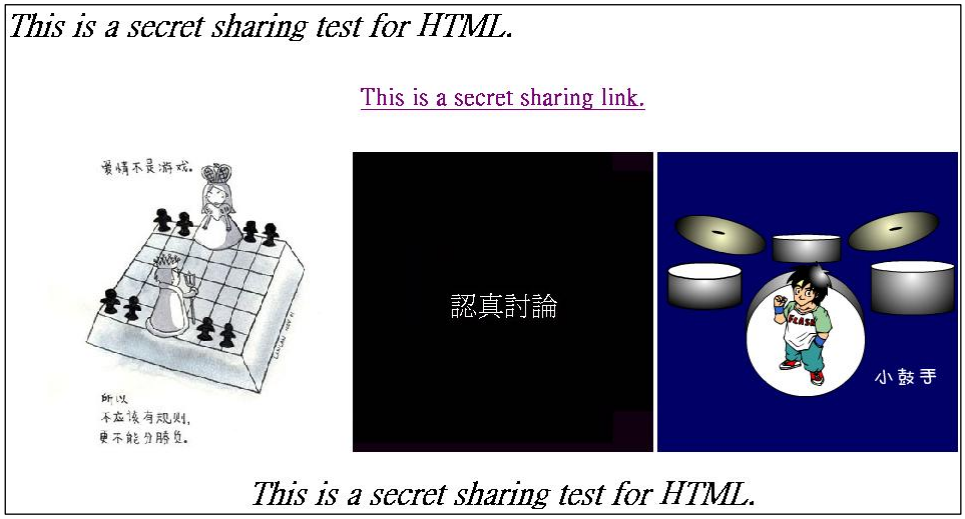


Figure 5: Recovered secret HTML document.

http://www.cis.nctu.edu.tw/~gis91568/Multimedia/image03.jpg

(a)

http://www.cis.nctu.edu.tw/~gis91568/Multimedia/image03.cgi?2856C1B0F09906517F697DCDBC7A4B3647B39D48C398B0ACF6F714D477F77C811C8249ACF49C125BE87E08646D12129698ABD97AEC511606EC71F0C01ABEDF768792B9CB1B963CC0F7D50054CFA5A676F7103A54B575B6CECA0D1A2D54BC24EF8CA547877122020072F90D350AD708C1354055878F0DBC

(b)

http://www.cis.nctu.edu.tw/~gis91568/Multimedia/image03.cgi?2856C1B0F09906517F697DCDBC7A4B3647B39D48C398B0ACF6F714D477F77C811C8249ACF49C125BE87E08646D12129698ABD97AEC511606EC71F0C01ABEDF768792B9CB1B963CC0F7D50054CFA5A676F7103A54B575B6CECA0D1A2D54BC24EF8CA547877122020072F90D350AD708C1354055878F0DBC

(c)

Figure 6: Image components. (a) The secret image component; (b) the corresponding fake image component of the first HTML-type share; (c) the corresponding fake image component of the second HTML-type share.

REFERENCES

[1] A. Shamir, "How to share a secret," *Communications of the Association for Computing Machinery*, vol. 22, no. 11, pp 612-613, 1979.

[2] Y. Desmedt and Y. Frankel, "Threshold cryptosystem," *Advances in Cryptology --- CRYPTO'89*, pp 307-315, 1989.

[3] T. P. Pedersen, "A threshold cryptosystem without a trusted party," *Advances in Cryptology --- EUROCRYPT'91*, pp 522-526, 1991.

[4] C. S. Laih and L. Harn, "Generalized threshold cryptosystems," *Advances in Cryptology --- ASIACRYPT'91*, pp 159-169, 1991.

[5] D. R. Stinson, "An explication of secret sharing schemes," *Designs, Codes, and Cryptography*, vol. 2, pp 357-390, 1992.

[6] C. C. Chang and H. C. Lee, "A new generalized group-oriented cryptoscheme without trusted centers," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 5, pp 725-729, 1993.

[7] H. M. Sun and S. P. Shieh, "Construction of dynamic threshold schemes," *Electronics Letters*, vol. 30, no 24, pp. 2023-2024, 1994.

[8] C. H. Cooke, "An integer Optimization Problem with mixed algebraic and number-theoretic constraints," *Applied Mathematics Letters*, vol. 16, pp 635-638, 2003.

[9] M. Hillery et al., "Quantum secret sharing," *Physical Review*, vol. 59, no. 3, pp 1829-1834, 1999.

[10] M. Greferath and S. E. Schmidt, "Secret sharing on partially ordered sets," *1998 Proceedings. 1998 IEEE International Symposium on*, p 299, 16- 21 Aug., 1998.

[11] A. Beimel and Y. Ishai, "On the power of nonlinear secret-sharing," *16th annual IEEE Conference on 2001-Computational Complexity*, pp 188-202, 2001.

[12] C.-S Laih and Y.-C Lee, "V-fairness (*t*, *n*) secret sharing scheme," *IEE Proceedings-Computers and Digital Techniques*, vol. 144, issue 4, pp 245-248, July 1997.

[13] Okamoto, E., "Cryptosystems based on polynomials over finite fields," *Information Theory Workshop, 2002. Proceedings of the 2002 IEEE*, pp 74-77, 20-25 Oct. 2002.

[14] C. C. Lin, "A study on secret sharing with steganographic effects," *Doctor Thesis, Department of Computer and Information Science, National Chiao Tung University, Taiwan, Republic of China*, 2003.

[15] Y. H. Chang and W. H. Tsai, "A steganographic method for copyright protection of HTML documents," *Proc. of 2003 National Computer Symposium*, Taichung, Taiwan, Dec. 200

APPENDIX: Proof of the effect of Equation (2) on controlling the data magnitude of a share

Let s be a one-byte secret, n be the number of secret sharing participants, and a_i be a randomly selected integer from the set {0, 1, …, 255}, where $i = 1, 2, \ldots, n$. We want to prove that if the values of all a_k together satisfy Equation (2), then $f(a_k)$ for $k = 1$ through n will not be 256. Suppose that $f(a_j) = 256$ and all a_i satisfy Equation (2). We want to conduct the proof by contradiction. First, Equation (2) may be deduced in the following way:

$$\begin{aligned} & f(a_j) = 256 \\ \Leftrightarrow & \left(\sum_{i=1}^{n} (a_j - a_i) + s \right)_{\text{mod } 257} = 256 \\ \Leftrightarrow & \left(n \times a_j - \sum_{i=1}^{n} a_i + s \right)_{\text{mod } 257} = 256 \\ \Leftrightarrow & \sum_{i=1}^{n} a_i = \left(n \times a_j - 256 + s \right)_{\text{mod } 257}. \end{aligned}$$

Under the assumption that all a_i satisfy Equation (2), we get further the following:

$$\begin{aligned} & n \times a_j - 256 + s = \left[(n-1) \times 256 + s \right]_{\text{mod } 257} \\ \Leftrightarrow & \; n \times a_j = (n \times 256)_{\text{mod } 257} \\ \Leftrightarrow & \; a_j = 256_{\text{mod } 257} \\ \Leftrightarrow & \; a_j = 256. \end{aligned}$$

But the possible value of a_j ranges from 0 to 255, and so contradiction occurs. This completes the proof of the claim.