

A New Authenticable Steganographic Method via the Use of Numeric Data on Public Websites

Che-Wei Lee, Bay-Erl Lai

Abstract—A new steganographic method via the use of numeric data on public websites with a self-authentication capability is proposed. The proposed technique transforms a secret message into partial shares by Shamir's (k, n) -threshold secret sharing scheme with $n = k + 1$. The generated $k+1$ partial shares then are embedded into the numeric items to be disguised as part of the website's numeric content, yielding the stego numeric content. Afterward, a receiver links to the website and extracts every k shares among the $k+1$ ones from the stego numeric content to compute $k+1$ copies of the secret, and the phenomenon of value consistency of the computed $k+1$ copies is taken as an evidence to determine whether the extracted message is authentic or not, attaining the goal of self-authentication of the extracted secret message. Experimental results and discussions are provided to show the feasibility and effectiveness of the proposed method.

Keywords—Steganography, data hiding, secret authentication, secret sharing.

I. INTRODUCTION

THE development of steganographic techniques provides a way to transmit important information secretly. Such techniques hide secret messages in an undetectable way via the use of various types of digital data, such as images, audios, texts etc., to complete the covert communication. In steganography, a sender hides a secret message into a *cover file*, generating a *stego-file*; and a receiver extracts the hidden secret message from the stego-file to accomplish the mission of secret transmission.

In the literary, to hide secret messages into image files, methods in [1]-[3] replace the least significant bit of pixels with secret message data. However, high embedding rate would inevitably cause much distortion to the content of image files. Methods in [4]-[6] using pixel-value differencing which refers to calculating the difference between two pixels' values in each image block, and convert the difference value with a sub-stream of the secret message. On the other hand, some information hiding methods utilizing other cover media to transmit secret messages. Lee and Tsai [7] encode a secret message with some special ASCII codes, and embed the encoded result at the between-word or between-character location in the text part of a cover PDF file. Also, the method proposed by [8] conceals the secret data in Microsoft Office 2007 files, and provide an algorithm to detect hidden data in such a file. Gopalan [9] take audio signals as cover media to

achieve steganography. In the method, one bit in each of samples of a given cover utterance is altered in accordance with the secret data bits. Qazanfari et al. [10] recently proposed an improved LSB^{++} steganographic method which uses DCT coefficients histogram of jpeg images and select sensitive pixels to protect them from extra bit embedding. Though there have been many methods developed for steganography [11]-[14], these existing methods are hard to deal with a situation that an attacker may subtly destroy or modify the passing-by files between a monitored sender and a receiver, leading to the loss of accuracy of the hidden secret message. In such a situation, a receiver may be misled by the incorrect secret message. Thus is the inspiration of the idea proposed in our method.

In this study, we propose a new steganographic method which transforms a secret message into partial shares by using a $(k, k+1)$ -threshold secret sharing method. Then, we take numeric data as the cover medium to conceal the generated partial shares and use a public website, which shows these numeric data, as a new communication platform. Different from cover media of using multimedia data such as images, audios, etc., modifications caused by the embedment of a secret message made on numeric data is comparatively inconspicuous. Furthermore, we use a public website to present the stego numeric content and so a receiver only need to link to the public website when necessary to extract the secret message from the stego numeric content, decreasing the risk of stego-files being intercepted during transmission between two sides. Importantly, the proposed method utilize a $(k, k+1)$ -threshold method [15] which is based on the (k, n) -threshold secret sharing method [16] to develop a new authenticable steganographic method. The purpose of the $(k, k+1)$ -threshold method is to achieve a self-authentication capability by checking the value consistency of all $k+1$ computed results. If the phenomenon of value consistency exists, the extracted secret message is determined correct; otherwise, the extracted message is inauthentic. As an example, a sender chooses a website on which there is a table containing numeric data. Next, use the proposed method to convert a secret message into $k+1$ partial shares that are presented in the form of numbers and embed these partial shares into the numeric data listed in the table, yielding a stego-table on the website. Later, inform a receiver the website address and the receiver may easily retrieve and verify the secret message from the stego-table of the website. Fig. 1 shows flowcharts of the proposed method.

Che-Wei Lee is with the Department of Information Management, National Kaohsiung Marine University, Kaosiung 81157, Taiwan (phone: +886 913200067; fax: +886 7 3658452; e-mail: paradiseree@gmail.com).

Bay-Erl Lai is with the Department of Information Management, National Kaosiung Marine University, Kaosiung, 81157 Taiwan (e-mail: 1001241144@stu.nkmu.edu.tw).

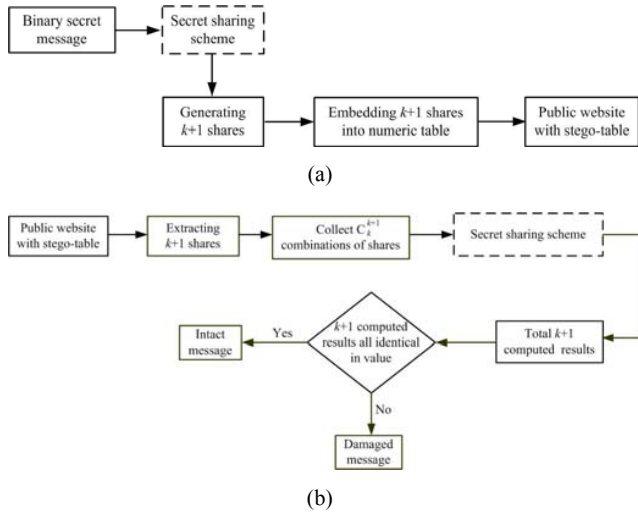


Fig. 1 Flowcharts of proposed authenticable steganographic method via the use of numeric data on a public website. (a) Generation of a public website with a stego-table containing numeric data. (b) Extraction and self-authentication of the secret message.

II. REVIEW OF SHAMIR'S METHOD FOR SECRET SHARING

The proposed authenticable steganographic method is based on the (k, n) -threshold secret sharing scheme proposed by Shamir. By transforming the secret d into n shares which then are disseminated to n participants to keep; and with at least k of the n shares are collected, the original content of the secret will be recovered. The detailed Algorithm of the method is reviewed in the following.

Algorithm 1. (k, n) -threshold secret sharing.

Input: a secret d in the form of an integer, the number n of participants, and a threshold k not larger than n .

Output: n shares in the form of integers for the n participants to keep.

Steps:

1. Randomly choose a prime number p which is larger than the secret d .
2. Select $k - 1$ integer values c_1, c_2, \dots, c_{k-1} within the range of 0 through $p - 1$.
3. Select n distinct real values x_1, x_2, \dots, x_n .
4. Generate n equations by following $(k-1)$ -degree polynomial to compute n function values $F(x_i)$:

$$F(x_i) = (d + c_1x_i + c_2x_i^2 + \dots + c_{k-1}x_i^{k-1})_{\text{mod } p} \quad (1)$$

where $i = 1, 2, \dots, n$.

5. Deliver the 2-tuple $(x_i, F(x_i))$ as a share to the i th participant, where $i = 1, 2, \dots, n$.

In (1), there are k coefficients consisting of d and c_1 through c_{k-1} and it is required to collect at least k shares from the n participants to form k equations of the form of (1) to solve these k coefficients in order to retrieve the secret d . This is the reason why the Shamir method is called (k, n) -threshold secret sharing method. The following algorithm is a description of the process for secret recovery.

Algorithm 2. Secret recovery

Input: collect k shares from the n participants and the prime number p which was chosen in Step 1 of Algorithm 1.

Output: the secret d hidden in the shares and other coefficients c_1 through c_{k-1} in the equation described in Algorithm 1.

Steps:

1. Collect k shares from n , $(x_1, F(x_1)), (x_2, F(x_2)), \dots, (x_k, F(x_k))$ to set up the following equations:

$$F(x_j) = (d + c_1x_j + c_2x_j^2 + \dots + c_{k-1}x_j^{k-1})_{\text{mod } p} \quad (2)$$

where $j = 1, 2, \dots, k$.

2. Solve the k equations above by Lagrange's interpolation to obtain the secret value d as follows:

$$d = (-1)^{k-1} \left[F(x_1) \frac{x_2x_3 \dots x_k}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_k)} + F(x_2) \frac{x_1x_3 \dots x_k}{(x_2 - x_1)(x_2 - x_3) \dots (x_2 - x_k)} + \dots + F(x_k) \frac{x_1x_2 \dots x_{k-1}}{(x_k - x_1)(x_k - x_2) \dots (x_k - x_{k-1})} \right]_{\text{mod } p} \quad (3)$$

3. Compute the values c_1 through c_{k-1} by expanding the following equation and compare the result with (2) in Step 1 while regarding the variable x in the equality below to be x_j in (2):

$$F(x) = \left[F(x_1) \frac{(x - x_2)(x - x_3) \dots (x - x_k)}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_k)} + F(x_2) \frac{(x - x_1)(x - x_3) \dots (x - x_k)}{(x_2 - x_1)(x_2 - x_3) \dots (x_2 - x_k)} + \dots + F(x_k) \frac{(x - x_1)(x - x_2) \dots (x - x_{k-1})}{(x_k - x_1)(x_k - x_2) \dots (x_k - x_{k-1})} \right]_{\text{mod } p} \quad (4)$$

III. PROPOSED STEGANOGRAPHIC METHOD BY USING NUMERIC DATA ON A PUBLIC WEBSITE

In the proposed method, the sender side first prepares numeric data on a website to be the cover medium. Here we use a table W containing numeric data as an example, but not be restricted to be so, to illustrate the proposed idea. The numeric data shown in the table are used to disguise generated secret shares. Next, we divide a secret message M into several segments and take them as input of (k, n) -threshold secret sharing with appropriate parameters to generate secret shares. Then use a secret key K to randomly choose the numeric items in W , and replace the chosen numeric items with shares to generate a stego-table W' . A detailed algorithm describing processes of proposed method is presented in the following.

Algorithm 3. Generation of a stego-table with numeric data on a website.

Input: a binary secret message M divided into m -bit segments, a website with a cover table of numeric data W , a secret key K , and three pre-selected integers k, n ($n = k + 1$), and m .

Output: a stego-table W' on a website.

Steps:

Stage 1 – share generation.

- Step 1. Select a table with numeric data on a public website prepared in advance.
- Step 2. Choose the appropriate prime number p whose value is larger than $2^m - 1$.
- Step 3. Take sequentially m bits from M to form segments and transform each segment into a partial share by performing the steps below:

- 3.1 Transform the k segments into integers and take the corresponding results to be values of $d, c_1, c_2, \dots, c_{k-1}$, respectively, where k is the number of segments.
- 3.2 Set x_1 through x_n to be the integers 1 through n , respectively, where $n = k + 1$.
- 3.3 Use the following $(k - 1)$ -degree polynomial to compute n partial shares $F(x_i)$:

$$F(x_i) = (d + c_1x_i + c_2x_i^2 + \dots + c_{k-1}x_i^{k-1})_{\text{mod } p} \quad (5)$$

Where $i = 1, 2, \dots, n$.

- 3.4 Save all $F(x_i)$ in order into a *partial-share* set F_{ps} .

Stage 2 – partial share embedding.

- Step 4. Use a secret key K to randomly select a numeric item I in W , and replace I with a partial share $F(x_i)$ selected sequentially from F_{ps} .
- Step 5. If partial shares in F_{ps} are all selected, take the final W as the output W' , otherwise, go to Step 4.

In the next Algorithm below for secret extraction, the receiver side links to the website to access the stego-table W' and perform the following steps to retrieve and authenticate the hidden secret message.

Algorithm 4. Secret message recovery and self-authentication.

Input: a stego-table W' ; the prime number p , the secret key K used in Algorithm 3, three integers $k, n (=k+1)$ and m .

Output: a secret message M hidden in W' , and a report about the authenticity of M .

Steps:

Stage 1 – message recovery.

- Step 1. Use the secret key K to select numeric items in W' whose values are presumably partial share $F(x_i)$ embedded by Algorithm 3; and put these selected items sequentially into a set F_{ps} as a partial-share set.
- Step 2. Take out the n partial shares from F_{ps} , set their corresponding x values as 1 through n , respectively, and perform the following steps to obtain the secret message M .
- 2.1 Take every k partial shares F_1, F_2, \dots, F_k from n ones in F_{ps} , and set their corresponding x values as x_1, x_2, \dots, x_k . Then, perform the following steps to obtain total $n = k + 1$ sets of values of d and c_1 through c_{k-1} .
- 2.1.1 Use the k shares $(x_1, F_1), (x_2, F_2), \dots, (x_k, F_k)$ to set up the following equations:

$$F_j = F(x_j) = (d + c_1x_j + c_2x_j^2 + \dots + c_{k-1}x_j^{k-1})_{\text{mod } p} \quad (6)$$

where $j = 1, 2, \dots, k$.

- 2.1.2 Expanding the following equality to compute the values d and c_1 through c_{k-1} , and comparing the result with (6) in Step 2.1.1 above while regarding the variable x in the equality below to be x_j in (6):

$$F(x) = \left[F(x_1) \frac{(x-x_2)(x-x_3)\dots(x-x_k)}{(x_1-x_2)(x_1-x_3)\dots(x_1-x_k)} + F(x_2) \frac{(x-x_1)(x-x_3)\dots(x-x_k)}{(x_2-x_1)(x_2-x_3)\dots(x_2-x_k)} + \dots + F(x_k) \frac{(x-x_1)(x-x_2)\dots(x-x_{k-1})}{(x_k-x_1)(x_k-x_2)\dots(x_k-x_{k-1})} \right]_{\text{mod } p} \quad (7)$$

Stage 2 – self-authentication of computed secret message.

- Step 3. Take the n sets coefficient values of d and c_1 through c_{k-1} and perform the following steps to gain the message content.

- 3.1 In order transform the coefficients values of d and c_1 through c_{k-1} into m -bit binary segment, and concatenate them to obtain a message M .

- Step 4. If all the n sets of coefficients are identical to each other in value, regard M as an authentic message; else, regard M as having been damaged.

IV. DISCUSSIONS ON THE PROPOSED METHOD

To the best of our knowledge, the proposed method is the first steganographic method using the numeric data presented on a public website to be the cover media for transmitting a secret message and, meanwhile, having the capability of self-authenticating the extracted message. Besides, some other important merits of the proposed method are further described as follows.

- (1) Developing a new way of communicating a stego-medium – Different from the conventional way of sending a stego-file to a receiver through network, the proposed method uses a public website as a communication platform to transmit the concealed secret and so a receiver only need to link to the website when necessary to extract the secret message. Such cloud-like secret transmission means a stego-file need not to be sent out to a receiver so that a sender no more has to worry the security of the stego-file during transmission. In such a way, not only the security of the secret message is assured but the variety of the secret communication in steganography is attained.
- (2) Avoiding conventional interception attacks – Attackers would monitor the network linking between a sender and a receiver to intercept the transmitted stego-file for malicious objectives such as cracking the secret message or modifying the content of the stego-file. However, in the proposed method, there is no direct connection between a sender and a receiver, which means that the secret communication is hard to be detected for attackers, avoiding the conventional interception attacks.
- (3) Having a capability detecting malicious alteration – By checking the value consistency of computed $k+1$ sets of coefficients, the proposed method is able to verify the correctness of extracted message. In other words, the proposed method is capable of disclosing the malicious alteration having been occurring to the stego-table on the website and so the corresponding remedial measures can be adopted.

V. EXPERIMENTAL RESULTS

An experiment using a cover table which contains numeric data about human statistics on a website has been conducted to test the proposed method. The website presents the content of the cover table is shown in Fig. 2. In this experiment, we embed the partial shares transformed from a secret message into last three numbers of the item “Home Number” shown in the cover table of Fig. 2. Note that the type of a cover table and its corresponding content used in this experiment are just an example for demonstrating the proposed method but they need not be restricted to be so.

Fig. 2 A cover table containing numeric data about human statistics on a public website

Then, we set values for the prime number p , m and k used in (5) of Algorithm 3 to be 997, 8 and 10, respectively. The value of the prime number p was set to be 997 because it makes the generated secret shares' value fall into a reasonable range of 0 through 996 that can be accommodated by the last three numbers of home numbers. The value $m = 8$ satisfies the restriction of $2^m - 1 = 255 < p$ mentioned in step 2 of Algorithm 3, which means the length of each segment of the input secret message M is 8 bits. Subsequently, each 8-bit message segment of M was transformed into an integer for use as one of the coefficients d , c_1 , c_2 , ..., c_{k-1} in (5) of Algorithm 3.

In the experiment, the input secret message M was taken to be "Go to 10 F" as shown in Fig. 3. The secret message composed of 10 characters was first transformed into binary string with $8 \times 10 = 80$ bits (8 bits per ASCII-coded character). The 80 bits are then divided into 10 message segments with each segment consisting of 8 bits as mentioned previously. And these 10 message segments will totally generate 11 partial shares which are then embedded randomly into the numeric items of home numbers in the cover table to yield a stego-table on a public website. As can be seen from Fig. 4, the last three numbers of phone numbers in the stego-table are replaced with values of partial shares. For example, the phone number of Kelvin Norman originally is 527-7917 shown in Fig. 2, but now, in the stego-table, is 527-7123. Additionally, note that in this case the value of k is 10 because there are 10 message segments as mentioned previously and so the value of n is $k + 1 = 11$.

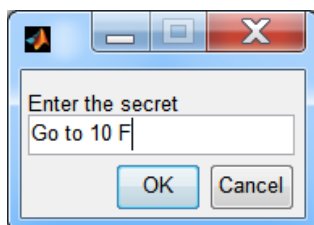


Fig. 3 A dialog for inputting secret message

Fig. 4 A stego-table with yellow regions where partial shares locate

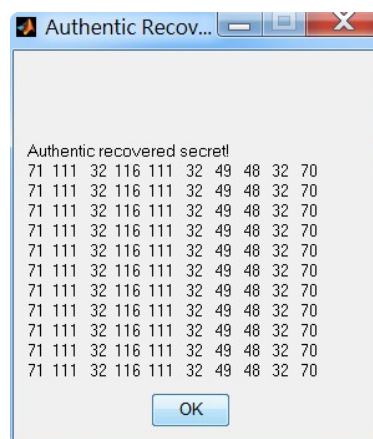


Fig. 5 A dialog shows the value consistency of computed 11 sets of coefficients extracted from the untouched stego-file

With the public website, a receiver can link to it for collecting the partial shares to retrieve the concealed secret message. Furthermore, the correctness of the retrieved secret message can be verified by applying Algorithm 4. Fig. 5 shows the phenomenon of value consistency existing in the computed 11 sets of coefficients. Therefore, as shown in Fig. 6, the recovered secret message transformed from computed coefficients that essentially are ASCII codes mentioned previously is intact and authentic. On the contrary, if the content of the stego-table suffers maliciously subtly modifications, the phenomenon of value consistency will not exist in the computed 11 sets of coefficients. Fig. 7 shows the attacked content of the stego-table in which the modified home numbers are marked in red. As can be observed, the modification attack destroyed four partial shares located in yellow regions. Therefore, after performing Algorithm 4, the computed 11 sets of coefficients are not identical to each other in value as shown in Fig. 8. Accordingly, a receiver will know the extracted secret is unreliable. In Fig. 9, the result of

extracted inauthentic secret is marked with question marks for warning.

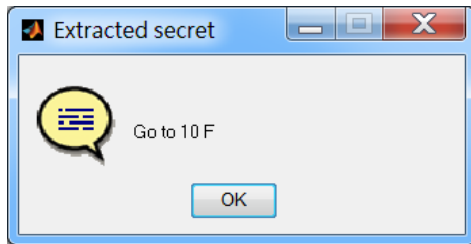


Fig. 6 A dialog with extracted secret message

	Name	Age	Gender	Day of Birth	Home Number	Postal Number
1	Kelvin Norman	9	M	2/10	837-1275	900
2	Christine Patel	23	M	7/7	827-3733	803
3	Rachel Vine	43	F	9/19	450-7214	800
4	Debra Jones	1	M	5/23	754-6916	700
5	Olivia Apple	12	F	6/15	804-6813	211
6	Lila Field	44	F	12/12	749-1027	367
7	Bentham Goodall	69	M	4/12	459-6122	272
8	Daryl Vine	78	M	7/19	525-4725	745
9	Jay Dickinson	35	M	8/8	934-1264	814
10	Lorraine Dennis	68	M	11/23	623-8286	701
11	Alfonso Apple	90	F	5/20	427-5129	600
12	Malachi Rose	33	F	2/21	285-8209	890
13	April Apple	46	F	1/30	612-5589	613
14	Madison Apple	66	F	4/4	222-9227	315
15	Bartholomew Edwards	73	M	2/28	390-6269	540
16	Gina Clark	23	M	1/19	836-9763	368
17	McDonald Dennis	11	M	9/30	233-6822	235
18	Nettie Brown	9	F	8/31	452-8109	300
19	Dafne Bell	5	F	5/5	525-6692	116
20	Ima Holloway	99	F	9/23	287-4472	817
21	Meredith Castle	55	F	8/1	394-4142	100
22	Francisco Hughes	58	M	9/5	526-7672	290
23	Katherine Fleming	98	F	10/31	826-9792	227
24	Ginny Patel	82	M	3/3	553-9632	203

Fig. 7 An attacked stego-table with modified home numbers marked in red

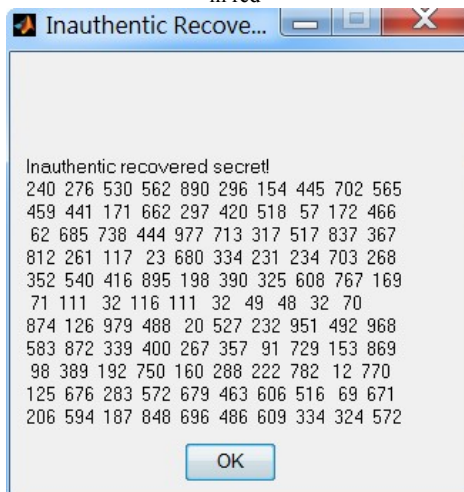


Fig. 8 A dialog shows the value inconsistency of computed 11 sets of coefficients extracted for the attacked stego-file

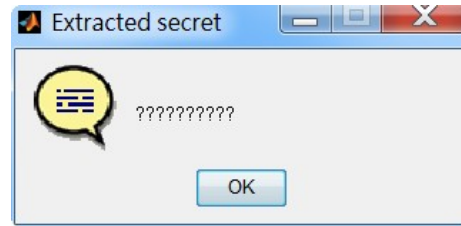


Fig. 9 A dialog with inauthentic secret message marked with question marks for warning

VI. CONCLUSION

A new authenticable steganographic method via the use of numeric data on public websites is proposed. By applying a transformed Shamir's ($k, k + 1$) secret sharing scheme, the segments of a secret message are converted into partial shares and embedded into numeric data presented on a public website, yielding a camouflage effect with the capability of self-authentication. To achieve the self-authentication capability, the authenticity of the secret message extracted from the stego-numeric-content can be verified by checking the value consistency of the computed results coming from all $k+1$ combinations of k shares out of $k+1$ ones. On the other hand, the proposed method using public websites as a new communication platform may diminish the suspicion of attackers, because there is no file transmission between senders and receivers. Experimental results have demonstrated the feasibility and effectiveness of the proposed method.

REFERENCES

- [1] Shailender Gupta, Ankur Goyal, Bharat Bhushan, "Information Hiding Using Least Significant Bit Steganography and Cryptography," *International Journal of Modern Education and Computer Science*, vol. 4, no. 6, pp. 27-34, 2012.
- [2] Khan, M. K., Naseem, M., Hussain, I. M. and Ajmal, A., "Distributed Least Significant Bit technique for data hiding in images," *2011 IEEE 14th International Multitopic Conference (INMIC)*, pp. 149-154, 2011.
- [3] Darabkh, K. A., Jafar, I. F., Al-Zubi, R. T., and Hawa, M., "An improved image least significant bit replacement method," *In Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014 37th International Convention on. IEEE. pp. 1182-1186, 2014.
- [4] Da-Chun Wu and Wen-Hsiang Tsai, "A steganographic method for images by pixel-value differencing", *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1613-1626, 2003.
- [5] Chung-Ming Wang, Nan-I Wu, Chwei-Shyong Tsai and Min-Shiang Hwang, "A high quality steganographic method with pixel-value differencing and modulus function," *Journal of Systems and Software*, vol. 81, no. 1, pp. 150-158, 2008.
- [6] Zhang, Xinpeng, and Shuozhong Wang. "Vulnerability of pixel-value differencing steganography to histogram analysis and modification for enhanced security." *Pattern Recognition Letters*, vol. 25, no. 3, pp. 331-339, 2004.
- [7] Lee, I-Shi, and Wen-Hsiang Tsai. "A new approach to covert communication via PDF files." *Signal Processing*, vol. 90, no. 2, pp. 557-565, 2010.
- [8] Park, Bora, Jungheum Park, and Sangjin Lee. "Data concealment and detection in Microsoft Office 2007 files." *Digital Investigation*, vol. 5, no. 3-4, pp. 104-114, 2009.
- [9] Gopalan, Kaliappan. "Audio steganography using bit modification." *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on. IEEE*, vol. 1, pp. I-629-32, 2003.
- [10] Qazanfari, Kazem, and Reza Safabakhsh. "A new steganography method which preserves histogram: Generalization of LSB++." *Information Sciences*, vol. 277, pp. 90-101, 2014.

- [11] Lee, Yeuan-Kuen, and Ling-Hwei Chen. "An adaptive image steganographic model based on minimum-error lsb replacement." *Nineth National Conference on Information Security*. pp. 8-15, 1999.
- [12] Chang, Chin-Chen, and Hsien-Wen Tseng. "A steganographic method for digital images using side match." *Pattern Recognition Letters*, vol. 25, no.12, pp. 1431-1437, 2004.
- [13] Holub, Vojtěch, and Jessica Fridrich. "Digital image steganography using universal distortion." *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, pp. 59-68, 2013.
- [14] Xinpeng Zhang. "Reversible data hiding in encrypted image." *Signal Processing Letters, IEEE*, vol. 18, no. 4, pp. 255-258, 2011.
- [15] C. W. Lee and W. H. Tsai, "A Covert Communication Method via Spreadsheets by Secret Sharing with a Self-Authentication Capability," *Journal of Systems and Software*, vol. 86, no. 2, pp. 324-334, 2013.
- [16] A. Shamir, "How to share a secret," *Communication of ACM*, vol. 22, pp. 612-613, 1979.