The Journal of Systems and Software 86 (2013) 324–334

Contents lists available at SciVerse ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

A covert communication method via spreadsheets by secret sharing with a self-authentication capability☆

Che-Wei Lee[a,1], Wen-Hsiang Tsai[a,b,*]

[a] Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan
[b] Department of Information Communication, Asia University, Taichung 41354, Taiwan

ARTICLE INFO

Article history:
Received 3 March 2012
Received in revised form 18 July 2012
Accepted 18 August 2012
Available online 30 August 2012

Keywords:
Covert communication
Secret sharing
Information hiding
Self-authentication
Spreadsheet

ABSTRACT

A new covert communication method with a self-authentication capability for secret data hiding in spreadsheets using the information sharing technique is proposed. At the sender site, a secret message is transformed into shares by Shamir's (k, n)-threshold secret sharing scheme with $n = k + 1$, and the generated $k + 1$ shares are embedded into the number items in a spreadsheet as if they are part of the spreadsheet content. And at the receiver site, every k shares among the $k + 1$ ones then are extracted from the stego-spreadsheet to recover $k + 1$ copies of the secret, and the consistency of the $k + 1$ copies in value is checked to determine whether the embedded shares are intact or not, achieving a new type of blind self-authentication of the embedded secret. By dividing the secret message into segments and applying to each segment the secret sharing scheme, the integrity and fidelity of the hidden secret message can be verified, achieving a covert communication process with the double functions of information hiding and self-authentication. Experimental results and discussions on data embedding capacity, authentication precision, and steganalysis issues are also included to show the feasibility of the proposed method.

1. Introduction

Covert communication is a technique of concealing secret information into a *cover medium* in an imperceptible way or with a camouflage effect such that only a sender and an intended receiver know the existence of the hidden data in the resulting *stego-medium*. In the literature, emphases were put on the use of multimedia like images, videos, and audios (Wu et al., 1999; Gopalan et al., 2003; Chae and Manjunath, 1999; Cheddad et al., 2010) because these media in general provide larger embeddable spaces and cause less suspicion due to their wide distributions. And weaknesses existing in human beings' visual capabilities are often exploited to design effective covert communication methods. For example, the methods proposed in Bender et al. (1996), Wu and Tsai (2003), and Yang et al. (2008) replace the least-significant bits of pixels in cover images to embed information, and that of Fridrich and Du (2000) uses the parities of palette colors, composed by similar colors, to represent hidden message bits.

In addition to methods developed for multimedia, several others (Brassil and Maxemchuk, 1999; Lee and Tsai, 2010a,b; Zhong et al., 2007; Liu and Tsai, 2007) used cover media of text, PDF, or Word documents for covert communication. In Brassil and Maxemchuk (1999), data are embedded by slightly adjusting the lines, tabs, or characters in text files. Lee and Tsai (2010a,b) used special ASCII codes in PDF files to embed data between characters. Liu and Tsai (2007) made use of the change tracking function in Microsoft Word to embed data imperceptibly by a document degeneration technique.

In this study, we propose a new covert communication method which applies Shamir's (k, n)-threshold secret sharing scheme (Shamir, 1979) with $n = k + 1$ to a given secret item to yield $k + 1$ shares, and the generated $k + 1$ shares are embedded into the number items in a spreadsheet as if they are part of the spreadsheet content. The purpose of transforming the secret data into secret shares by the $(k, k + 1)$-threshold secret sharing scheme is *not* to enforce robustness, but to yield a blind self-authentication capability for the embedded secret. Conventionally, the concept of (k, n)-threshold secret sharing is applied to provide destruction-tolerant capabilities. That is, any k shares collected from n ones may be processed to reveal the shared secret even though up to $(n - k)$ shares are destroyed. But in the proposed method, the scheme of $(k, k + 1)$-threshold secret sharing is developed for the first time

☆ This work is supported financially by the National Science Council, Taiwan, ROC under Project No. 99-2631-H-009-001.

* Corresponding author at: Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan.
Tel.: +886 3 5728368; fax: +886 3 5734935.
E-mail addresses: paradiserlee@gmail.com (C.-W. Lee), whtsai@cis.nctu.edu.tw (W.-H. Tsai).
[1] Tel.: +886 3 5728368; fax: +886 3 5734935.

0164-1212/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2012.08.048

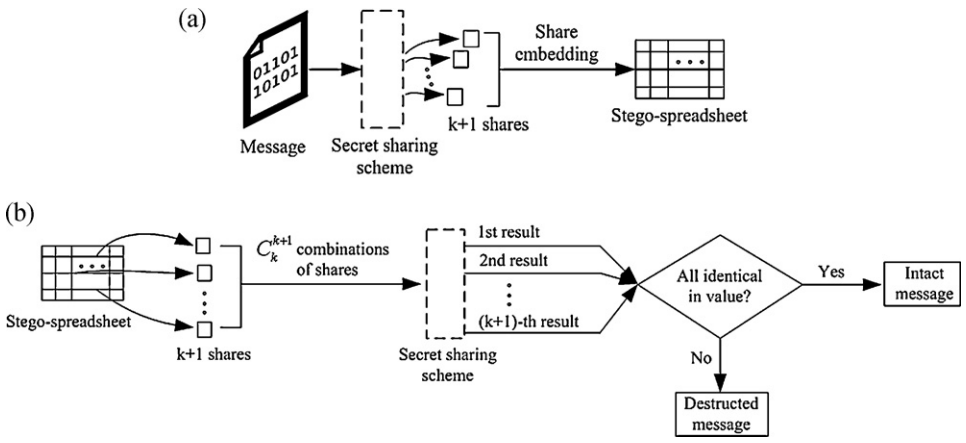

Fig. 1. Illustration of proposed covert communication method via spreadsheets by secret sharing. (a) Generation of a stego-spreadsheet. (b) Self-authentication of the extracted message.

to provide instead a self-authentication capability by checking the *value-consistency* of $k+1$ results coming from all $k+1$ combinations to determine whether the extracted secret is intact or not. That is, only when the results computed from any k shares collected from $k+1$ shares are *all identical* in value can the extracted secret be decided to be intact. Fig. 1 illustrates these core ideas of the proposed method.

Moreover, to conceal the presence of hidden data, secret shares are spread throughout the cover spreadsheet in a *sparsely* fashion. And a spreadsheet containing numeral items with a *high scatter level* is more suitable to be used as a cover spreadsheet for better concealment. Merits of the proposed method include the following. (1) A receiver can confirm the correctness of the extracted secret message. (2) Compared with some methods using hash codes or parity bits as redundant data to ensure the authenticity of retrieved data, only a minor redundancy, i.e. the $(k+1)$-th share, is needed in the proposed method. (3) By adaptively choosing involved parameters, i.e. the value of p, used in the polynomial of Shamir's method for the selected spreadsheet, the numerical items' values generated by the method will fall into a reasonable range of values, arousing little suspicion during covert communication. (4) Using spreadsheets as cover media, the proposed method is free from *unintentional* destruction of hidden data like data compression during the secret transmission or data keeping process, in contrast with cover media like images or videos which are often compressed *ignorantly* in such a process. Two examples of such documents, Microsoft Excel and Google Docs, are shown in Fig. 2.

The remainder of this paper is organized as follows. In Section 2, the Shamir method on which the proposed method is based is reviewed first. In Section 3, the details of the proposed method, including secret message embedding, secret message extraction, and self-authentication of the extracted message, are described. In Section 4, discussions on related issues about the proposed method are given. Experimental results are presented in Section 5, followed by conclusions in Section 6.

2. Review of Shamir's method for secret sharing

In the (k, n)-threshold secret sharing scheme proposed by Shamir (1979) with $k \le n$, a secret d in the form of an integer is transformed into shares which then are distributed to n participants to keep; and as long as up to k of the n shares can be collected, the original secret can be recovered. The detail of the scheme may be described as two algorithms 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 n participants to keep.

Steps.

1. Choose randomly a prime number p which is larger than the secret d.
2. Select $k-1$ integer values $c_1, c_2, \ldots, c_{k-1}$ within the range of 0 through $p-1$.
3. Select n distinct real values for the variables $x_1, x_2, \ldots, x_n$.
4. Use the following $(k-1)$-degree polynomial to compute n function values $F(x_i)$, called *partial shares*:

$$F(x_i) = (d + c_1 x_i + c_2 x_i^2 + \cdots + c_{k-1} x_i^{k-1})_{\bmod p}, \quad (1)$$

for $i = 1, 2, \ldots, n$.

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

Since there are k coefficients, including d and c_1 through c_{k-1}, in (1) above, it is necessary 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 recover the secret d. This explains the term, *threshold*, for k and the name, (k, n)-*threshold*, for the Shamir method. Below is a description of the equation-solving process for secret recovery.

Algorithm 2. Secret recovery.

Input: k shares collected from the n participants where k is the threshold mentioned in Algorithm 1; and the prime number p which was chosen in Step 1 of Algorithm 1.

Output: the secret d hidden in the shares and the coefficients c_i used in the equations described by (1) in Algorithm 1, where $i = 1, 2, \ldots, k-1$.

Steps.

1. Use the k shares $(x_1, F(x_1)), (x_2, F(x_2)), \ldots, (x_k, F(x_k))$ to set up the following equations:

$$F(x_j) = (d + c_1 x_j + c_2 x_j^2 + \cdots + c_{k-1} x_j^{k-1})_{\bmod p}, \quad (2)$$

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