Discussion on Steganographic Methods from the Perspective of E-Voting Implementation

Arman Avetisyan, Mariam Haroutunian and Parandzem Hakobyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
email: armanavetisyan1997@gmail.com, armar@sci.am, par_h@iiap.sci.am

***Abstract*—In this paper, a classification survey of steganographic algorithms existing in the literature is provided based on the embedding methods. The research is conducted to find out the advantages and issues of techniques in terms of applications in e-voting systems.**

***Keywords*— Steganography methods, embedding algorithms, hidden data transfer.**

I. Introduction

The aim of steganography is to convey secret messages by embedding them within other data to hide the fact of communication. There are 3 main types of steganography algorithms in the literature [1], shown in Fig.1.

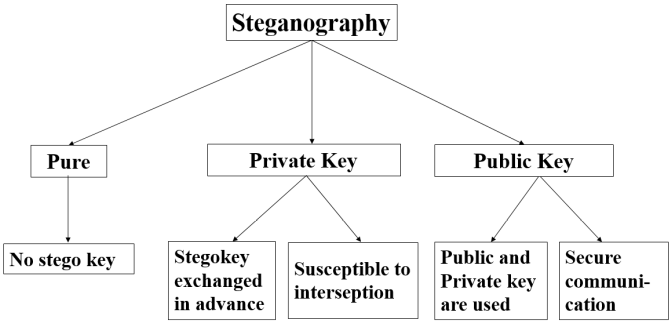

Fig. 1. Steganography Algorithms

Pure steganography algorithms do not use stegokeys and are not secure, if there is an observer over the channel, they can easily extract the hidden message. These algorithms are disregarded nowadays due to their impracticality.

Private key algorithms use only the private stegokey, which is exchanged in advance. They are less secure than public key algorithms, but give more freedom in terms of implementation and choice of communication channels.

Public key algorithms are the most secure. They use public and private keys for both correspondents to make communication as secure as possible.

Private key algorithms are generally a good middle ground when it comes to security and implementation, if additional security is required, it is recommended that additional encryption algorithms be used to protect data during transmission.

Examples of steganographic models that use private key algorithms can be seen in [2] - [5].

Based on the various steganographic models proposed in the literature, a classification of six steganographic methods (see Fig. 2) is proposed in [6].

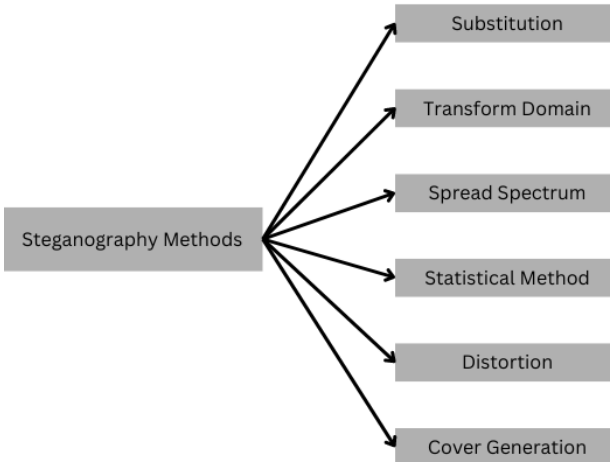

Fig. 2. Steganography Methods

- **Substitution** method replaces redundant parts of cover-text with a hidden message before sending.
- **Transform domain** method embeds secret information in the data transformation space.
- **Spread spectrum** method uses a communication method to propagate data over a channel in a certain way.
- **Statistical** method modifies the statistical properties of the covertext and the receiver uses hypothesis testing in the extraction process.
- **Distortion** method stores information by distorting the data, and the receiver compares it with the original.
- **Cover Generation** method creates an entirely new cover/copy of the data to hide the information.

The development of reliable and secure e-voting systems remains relevant in the world due to its complexity. Many models have been suggested and used but, unfortunately, a lot of vulnerabilities have been found in a comparative analysis of the most widely used e-voting approaches in [7]. Based on these vulnerabilities, a model was suggested in [8] that addresses them, but one of the most important issues still remains - secure data transfer from the storage server that keeps all the votes, to the counting server that should count them anonymously. Modern systems rely on physical copies of data

and are transferred manually, which is a major security issue and, thus, the usage of steganography methods is explored for that step. The main concern is that the channel between servers may be vulnerable to attacks and if an adversary can send unreliable data to the counting server, the entire electoral process will be in doubt.

We will discuss the advantages and disadvantages of each method in terms of possible e-voting applications.

II. Review of steganographic methods from e-voting standpoint

A. Substitution Method

The substitution method replaces redundant or unneeded bits of data to embed a secret message in them. Most popular steganography tools nowadays use the Least-Significant Bit (LSB) algorithm. There are many wasted bits in digital data such as images or audio files, the LSB algorithm takes advantage of these bits and hides the message. It is widely used because the algorithms are fast and easy to implement. It works very well with both grayscale and color images. The drawback of using this method is that depending on the altered bit, it might have apparent consequences on the data, thus compromising the whole purpose of steganography. Also, if the image is modified by a third party, even if it is rotated, the extraction algorithm will not be able to find the bits that have been modified.

A number of techniques have been developed using the substitution method. Two component-based LSB algorithm integrated with AES is suggested by M. Juneja and P. Singh Sandhu in [9]. P. Thiyagarajan et al. used 3D models resistant to uniform affine transformations [10]. Sh. Ahmed Laskar and K. Hemachandran implemented data embedding in the red plane of the selected image using PRNG [11].

B. Transform Domain Method

The transform domain technique is generally used when digital data need to be converted before being sent. The best example of data transformation is compression. The JPEG format is one of the most popular formats for sending images. It is interesting because the compression takes place when one closes the file. That is done to get rid of excess data and all the excess bits, that is why the substitution technique does not work in this case. Instead of substitution, the transformation itself can be used to hide information, when a JPEG is closed and thus compressed, the specific change that it goes through can be used to send and extract the secret message. This method is evidently harder to use in broader situations, since the format of the data sent must be quite specific for the algorithm to work. Nevertheless, the transform domain is quite interesting, because the data must undergo transformations again when received, making it difficult for a third party to detect the stegodata in the communication channel.

Integer wavelet transform technique was suggested in [12] for comparing embedding in two different domains. This technique was also used to hide multiple secret images in a cover image (see [13]).

Using image steganography, covert communication channels were established in [14] with additional encryption.

C. Spread Spectrum Method

The spread spectrum technique is one of the least popular methods because it imposes many restrictions on the communication channel [15]. We can differentiate two types of spread spectrum techniques:

- **Direct Sequence** method divides data to be transmitted into small pieces. Each of the pieces is assigned to a specific frequency in the channel using a predetermined spreading algorithm. The pieces have a redundant bit sequence code, which is used to embed the message. It also makes the pieces resistant to interference and helps recover data in case of damage.
- **Frequency Hopping** method first divides the bandwidth into many possible broadcast frequencies and then performs the direct sequence technique. This way the channel can be changed constantly but it takes a lot of power, so most of the time it is better to use only the direct sequence.

Generally, direct sequence techniques are impractical to use at present due to the need to have a channel with the ability to send and receive data at a specific frequency, such as radio transceivers, but other techniques hold up much better on a larger scale.

D. Statistical Method

The main difference of the statistical method is that a person can first figure out whether a hidden message has been embedded in the data or not, without using an extraction algorithm. A lot of data can be sent through a channel, and only a small percentage might have a hidden message in it, which is ideal when a public or potentially compromised channel is being used. For explanation, let's assume that the data is sent in text form. The sender has the choice of sending a Covertext (text without any hidden messages) or Stegotext (text with a hidden message embedded in it). For each of them, the sender should take data from a set distribution of either stegotexts or covertexts. The receiver can test the hypothesis on the received data and compare it with stegotext and covertext distributions, predicting whether the received text contains a hidden message or not. This creates a natural "cover" when used in a public channel because the observer cannot be sure which data are stegotexts. The most popular statistical method is called a "one-bit" method, when even a single bit alteration creates a statistical change that is significant enough to allow to conduct a hypothesis testing and make an accurate decision.

The implementation of the statistical method algorithm, depending on the choice of hypothesis testing technique, has been studied in a number of works. Two-stage optimal hypothesis testing for extraction and authentication was studied in [5]. Statistical techniques were suggested using syndrome trellis codes with additive distortion function in steganography and matrix embedding with wet paper codes in [16], [17] respectively. A reversible embedding scheme for

VQ-compressed images based on side matching and relocation using a location map was also considered from a statistical point of view [18].

E. Distortion Method

The distortion method creates certain changes in the data in order to hide information. The algorithm can alter the data immensely, and when the receiver compares the distorted data with the original, he can recover the message based on what alterations have been made. This method is a more "aggressive" version of substitution and allows more bits of the data to be changed since the sender doesn't care about keeping the likeness of original data. On the other hand, this method requires the receiver to have the original data, which is rather inconvenient to compare against, since the data essentially needs to be doubled, and distortions are obvious to an outside observer, potentially compromising the security of the model.

Distortion techniques were used with error-correcting codes in [19] and image blurring with sequential LSB embedding in [20].

F. Cover Generation Method

The cover generation technique is rather uncommon but unique. Usually, the message is hidden in the data itself, but using the cover generation method a cover over the data is created and its sole purpose is hiding the information. Spam mimicking is the most popular example of this method when a hidden message is sent along with the spam, which is generally ignored by the observer. This method is pretty costly though, because it requires the creation of a completely different dataset to work.

G. Discussion

Now, to see if we can use steganography methods for data transfer, we take a look at the type of data that is sent between servers. Assuming the voter data is securely entered into the storage server, it is then validated and cleared of the voter data, so that the votes remain anonymous even to the counting server. We are then left with a database of votes consisting of voter keys, which are used for validation and cannot be used to obtain the voter's identity (i.e., hash keys) and the vote itself, which is basically the candidate's index. The data does not contain any media data, it is dangerous to use substitution techniques to hide the candidate index in the voter key because this may corrupt the key and the counting server will be unable to verify the validity of the voter. We do not use data compression, we do not want to distort it, from all the methods described in the survey, the **statistical method** is best suited for the situation.

We have to send the data through a channel that might be prone to outside attacks, the data may be corrupted or the attacker may try to imitate the data and send fake votes to the counting server. A model that addresses the attacks by an active adversary is described in [5]. Using two stage hypothesis testing, we first decide if the data is valid and has a candidate index in it, and then we decide if it is an outside attacker trying to imitate valid votes. Using statistical methods (i.e., one-bit method), we can considerably simplify the first step without the need to use a steganographic extraction algorithm to find out if the data is a valid vote or, for example, noise. A model based on the statistical method and hypothesis testing gives the best approximation of the real-life implementation and is highly testable, the error probabilities can be calculated for both stages of data extraction giving a clear understanding of the reliability of the model.

III. Conclusion and future work

Most popular steganography methods are studied in terms of their viability when used for a secure electronic voting model. The statistical method was found to be most useful and practical. Furthermore, it is worth exploring the possibilities of creating a hybrid method on top of a statistical one when building an e-voting system. The modular e-voting system model will allow the usage of several steganography methods for added security.

Acknowledgment

The work was supported by the Science Committee of RA, within the framework of the research project № 21T-1B151.

References

[1] A. R. Remya and A. Sreekumar, "An inductive approach to the Knack of steganology", *International Journal of Computer Applications*, vol. 72, pp. 27-31, 2013.

[2] C. Cachin, "An information-theoretic model for steganography", *Information and Computation*, vol.192, pp. 41–56, 2004.

[3] M. Haroutunian, E. Haroutunian, P. Hakobyan and H. Mikayelyan, "Logarithmically asymptotically optimal testing of statistical hypotheses in steganography applications", *Proceedings of Intern. Conf. Collaborative Technologies and Data Science in Smart City Applications*, Logos Verlag Berlin, pp. 157-163, 2018.

[4] J. Shikata and T. Matsumoto, "Unconditionally secure steganography against active attacks," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2690-2705, June 2008

[5] M. Haroutunian, P. Hakobyan, A. Harutyunyan and A. Avetisyan, "Information-theoretic investigation of authenticated steganographic model in the presence of active adversary," *Proceedings of Intern. Conf. Collaborative Technologies and Data Science in Smart City Applications*, pp. 1-7, 2022.

[6] S. Katzenbeiser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Computer Security Series, Boston, London, 1999.

[7] A. Avetisyan, "Comparative analysis of modern e- voiting systems based on security criteria", *Proceedings of International Conference CSIT 2021*, Yerevan, Armenia, pp. 81-84, 2021.

[8] A. Avetisyan, "Electronic voting system essentials and problems", *Mathematical Problems of Computer Science*, vol. 57, pp. 39–46, 2022.

[9] M. Juneja and P. Singh Sandhu,"A new approach for information security using an improved steganography technique", *Journal of Information Processing Systems*, vol. 9, no. 3, pp. 405-424, 2013.

[10] P. Thiyagarajan, V.Natarajan, G.Aghila, V.Pranna Venkatesan and R.Anitha, "Pattern Based 3D Image Steganography", *3D Research center, Kwangwoon University and Springer, 3DR Express.*, pp.1-8, 2013.

[11] Sh. Ahmed Laskar and K. Hemachandran, "Steganography based on random pixel selection for efficient data hiding", *International Journal of Computer Engineering and Technology*, vol.4, no. 2, pp. 31-44, 2013.

[12] S. Hemalatha, U. Dinesh Acharya and A. Renuka, "Comparison of secure and high capacity color image steganography techniques in RGB and YCBCR domains", *International Journal of Advanced Information Technology*, vol.3, no.3, pp.1-9, 2013.

[13] S. Hemalatha, U. Dinesh Acharya, A. Renuka, P. Kamnath, "A secure and high capacity image steganography technique", *Signal & Image Processing – An International Journal*, vol.4, no.1, pp. 83-89, 2013.

[14] K. L.Haynes, "Using image steganography to establish covert communication channels", *International Journal of Computer Science and Information Security*, vol. 9, no. 9, pp. 1-7, 2011.

[15] L. M. Marvel, C. G. Boncelet and C. T. Retter, "Spread spectrum image steganography," *IEEE Transactions on Image Processing*, vol. 8, no. 8, pp. 1075-1083, Aug. 1999.

[16] T. Filler, J. Judas and J. Fridrich, "Minimizing Additive Distortion in Steganography using Syndrome Trellis Codes", *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp.920–935, 2011.

[17] J. Fridrich, M. Goljan and D. Soukal, "Wet paper codes with improved embedding efficiency ", *IEEE Transactions on Information Forensics and Security*, vol 1. no.1, pp. 102-110, 2006.

[18] Ch.-Ch. Chang and Ch.-Y. Lin, "Reversible steganography for VQ-compressed images using side matching and relocation ", *IEEE Transactions on Information Forensics and Security*, vol. 1, no.4, pp 493–501, 2006.

[19] M.B. Ould Medeni and E. Mamoun Souidi, "Steganography and error correcting codes", *International Journal of Computer Science and Information Security*, vol.8, no.8, pp. 147–149, 2010.

[20] D.P.Gaikwad and S.J.Wagh, "Colour image restoration for an effective steganography", *Imanager's Journal on Software Engineering*, vol.4, no.3, pp. 65-71, 2010.