On Complexity of Feature Descriptors and Elementary Cellular Automata in Steganography: A Proposal

Arezoo Abdollahi

September 24, 2017

Abstract

Information security is a trend in the technological era. To secure the transmission of secret data over the public network, various schemes have been presented. Steganography combined with cryptography can be one of the best choices for solving this problem. In this proposal, a new steganographic method based on making a cipher by applying image descriptors and Rule 30 of cellular automata are proposed. The proposed method is evaluated by objective analysis using various image quality assessment metrics. This approach is further analyzed concerning complexity measures.

1 Introduction

Preserving the privacy of personal information in various data management, dissemination, and mining applications have always been an issue. On the one hand, publishing data would help for data mining. On the other hand, by releasing data, the risk of disclosing sensitive information will be raised. Each record in such databases consists of three categories of information: identity attributes, quasi-identity attributes, and sensitive attributes. Therefore, individuals intend to protect their sensitive information from exposure to unauthorized parties via direct disclosure or indirect inferences [1]. A typical scenario of Privacy-Preserving Data Publishing is depicted in Figure 1, which shows the different phases of the data processing.

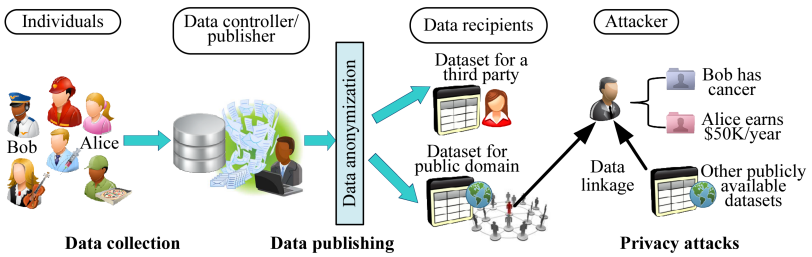

Figure 1: Overview of Privacy-Preserving Data Publishing [2]

While the scope of data privacy involves techniques such as cryptography, perturbation, auditing, and the like, this proposal particularly focuses on the problem of privacy protection using the data hiding techniques in an image, known as Steganography. This method has a distinct advantage over cryptography in which the intended secret message does not attract attention to itself as an object of scrutiny [3]. Indeed, the cryptography is the practice of protecting the contents of a message alone, whereas steganography is concerned with concealing the fact that a secret message is being sent as well as concealing the contents of the message. Our motivation is to combine the strength of digital image encryption with the advantage of steganography to hide the true information of the source in an image. Our focus would be on analyzing the complexity of different feature descriptors such as SIFT [4], SURF [5], LBP [6], and HOG [4] when they are fed to rule 30 of elementary cellular automata [7] to make an encrypted message. To do so, different complexity measures such as absolute complexity, relative complexity, entropy followed by morphological richness [8], Levenshtein distance [9], and peak signal-to-noise ratio (PSNR) [10] can be used.

2 Background

Information hiding involves providing security by obscurity for the participants of the dialogue, secrecy of the messages (steganography), or protection of the carrier (copyright marking). The roots of these methods stem from historic times, the need for sending messages that cannot be compromised in case of interception. Steganography is the science of invisible communication while steganalysis is the detection of steganography. Steganography means "covered writing" that hides the existence of the message itself. In today's world, various multimedia contents such as audio, text, video, image can act as cover media to carry secret information. Steganography, watermarking, and cryptography are closely related concepts. Steganography is hidden writing where as cryptography is secret writing. Digital watermarking is another branch which is similar to steganography in concept of data embedding, but there are several differences among them. A detailed comparison can be found in [3], [11].

Image steganography can be broadly classified into spatial domain, transform domain, spread spectrum and model based steganography. In spatial domain, secret message is embedded in pixel value directly whereas transform domain methods achieve embedding by first transforming the image from spatial to frequency domain. Spread spectrum steganography involves embedding in noise inherent to image acquisition process. Image restoration and error control techniques can be used while extracting the data at the decoder side. Model based steganography (statistics aware embedding) is based on statistical model of the cover image. In this category, before selecting the locations for data hiding in cover image, statistical global features of image are taken into account and then actual data embedding process is carried out accordingly. Thus, it provides additional layer of security to steganography.

Many steganographic tools based on low sign bit (LSB) substitution data hiding are available, e.g. StegHide, S tool, Stegnos etc. [12]. Adaptive LSB substitution methods work based on brightness, edges and texture masking of the host image to estimate the number k of LSBs for data hiding [13], loss

less generalized LSB data embedding [14], optimized LSB substitution using cat swarm strategy and genetic algorithm [15], data hiding based on histogram modification [16]. Ramaiya et al. developed a spatial domain steganography scheme based on S box mapping and secret key. The scheme is more secure as unintended recipient will not be able to extract secret message as information about mapping functions and secret key will not be available [17]. Simplicity and high perceptual efficiency are the advantages of these family methods. However, they suffer from inefficiency in terms of slight image manipulations such as scaling, rotation, cropping and addition of noise or lossy compression.

Digital image is a combination of low and high frequency components. The smooth and plane areas represent low frequency content whereas the edges and sharp transitions relate to the high frequency components. Low frequency regions are more sensitive as any change in them will be transparent to human visual system (HVS). Hence, hiding an equal amount of information in both high and low frequency regions makes the method unfeasible. Obtaining and analyzing the image in frequency domain, however, will greatly help to achieve efficient data embedding while it is less prone to attacks. Various image transforms that can be employed for data embedding include DCT, DWT, Haar transform, Hadamard transform, integer transform, contour let transform, DD DT DWT, Ridgelet transform, Ripplet transform and so forth. Mali et al. [18] use DCT coefficients and randomly spreads the embedded information all over the cover image. This method uses Image Adaptive Energy Thresholding (AET) Coding framework with Class Dependent Coding Scheme (CDCS) to make it robust against image compression, tampering, resizing, filtering and AWGN. It achieves minimum IQM variations, and a PSNR of 40 dB for 6190 bits of information. Chu et al. [19] uses similarities of DCT coefficients between the adjacent image blocks. In this manner, the method could well preserve image quality while the embedding distortion is spread within the image blocks. Huang et al. [20] utilize successive zero coefficients of the medium-high frequency components in each reconstructed block for 3-level 2-D DWT of a cover image which employs 9/7 wavelet filter in discrete wavelet transform (DWT). They offered PSNR of 31.41 dB for 36710 bits of hiding capacity and preserve the high quality of stego-image. Ghasemi et al. [21] embed data in 4×4 blocks of integer wavelet transform coefficients by using a mapping function based on genetic algorithm. After embedding the message, OPAP is applied. This method outperforms adaptive steganography technique based on wavelet transform in terms of PSNR and capacity, 39.94 dB and 50%, respectively.

The solution to the drawbacks of spatial domain steganography can be found with frequency domain approach. However, simplicity of spatial domain method certainly matters in design and cannot be ignored. To handle these issues, model based steganography is developed. P. Sallee [22] uses statistical properties of the cover medium to embed the secret message without altering any of cover's properties. Model based method works better by providing additional security layer as it is statistics aware, no steganography system is 100% secure. A popular adaptive method presented by Hioki [23] in which pixels of noisy blocks in an image is replaced with the block obtained by embedding data. Larger embedding capacity is one of the key advantages. Since modifications in the image are done by first analyzing the statistical parameters of the image, there will not be many issues with stego-image generated and will result in high quality.

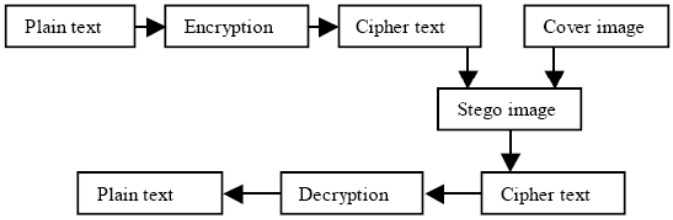

Figure 2: Overview of Privacy-Preserving Data Publishing [2]

3 Description of Proposed Research

With respect to the aforementioned advantages of model based steganography, we intended to propose an efficient method in this paradigm. The theme of model based steganography can be explained as below. Cover image is represented as a random variable X, and is divided in two parts, X_a that will not be modified during embedding and X_b, which is used to carry the secret message without altering the statistical properties of the cover. The embedded message is assumed to be a uniform random stream of bits. The embedded message is processed by an entropy decoder according to the conditional probability distribution $P_{X_b|X_a}(x_b|x_a)$. Its output is denoted by x'_b and forms together with X_a the stego message x'. At the decoder side, entropy encoder is used. The stego message x' is separated in x_a and x_b. Probability distribution $P_{X_b|X_a}(x'_b|x_a)$ is calculated to obtain x_b according to the model distribution and the encoder outputs the embedded message. A pictorial representation of the combined concept of cryptography and steganography is depicted in figure 2.

The procedure of the proposed research can be summarized as follows:

1. Four different feature descriptors which are SIFT, HOG, SURF, and LBP, are applied to the cover image. The most promising descriptors from each feature vector is then selected to make a raw key.

2. The key is passed into Rule 30 of elementary CA with a predefined iteration to make a chaotic cipher.

3. The message is encrypted with XOR the content by the generated key.

4. The encrypted data is hidden within the cover carrier.

5. For hiding message, a standard steganography method, such as embedding algorithm based on histogram-preserving data mapping which is proposed by Eggers et al. [24], is utilized.

6. The first step of the reverse procedure is done by extracting the message from the cover. Then the descriptors on which both sender and receiver were aggregated is extracted from the cover image and is passed to Rule 30 to make the key. Finally, the message is decrypted using the cipher.

Experiment would be done based on two scenarios. In the first scenario, the proposed method is compared to the state-of-the-art ones in terms of image quality assessment metrics. Security against attack, payload capacity, and

imperceptibility are other important metrics that will be reported. In the second scenario, however, we focus on complexity analysis which plays a vital role in algorithms in which the CA is employed. To reach this objective, different complexity measures such as absolute complexity, relative complexity, entropy followed by morphological richness, and Levenshtein distance can be used.

References

[1] Ting Wang and Ling Liu. From data privacy to location privacy. *Machine Learning in Cyber Trust*, pages 217–246, 2009.

[2] Vanessa Ayala-Rivera, Patrick McDonagh, Thomas Cerqueus, and Liam Murphy. A systematic comparison and evaluation of k-anonymization algorithms for practitioners. *Trans. Data Privacy*, 7(3):337–370, 2014.

[3] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt. Digital image steganography: Survey and analysis of current methods. *Signal processing*, 90(3):727–752, 2010.

[4] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[5] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[6] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[7] Stephen Wolfram. Cellular automata as models of complexity. *Nature*, 311(5985):419–424, 1984.

[8] Andrew Adamatzky and Genaro J Martinez. On generative morphological diversity of elementary cellular automata. *Kybernetes*, 39(1):72–82, 2010.

[9] Gonzalo Navarro. A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1):31–88, 2001.

[10] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.

[11] Mansi S Subhedar and Vijay H Mankar. Current status and key issues in image steganography: A survey. *Computer science review*, 13:95–113, 2014.

[12] Walter Bender, Daniel Gruhl, Norishige Morimoto, and Anthony Lu. Techniques for data hiding. *IBM systems journal*, 35(3.4):313–336, 1996.

[13] Hengfu Yang, Xingming Sun, and Guang Sun. A high-capacity image data hiding scheme using adaptive lsb substitution. *Radio Eng*, 18(4):509, 2009.

[14] Mehmet Utku Celik, Gaurav Sharma, A Murat Tekalp, and Eli Saber. Lossless generalized-lsb data embedding. *IEEE transactions on image processing*, 14(2):253–266, 2005.

[15] Zhi-Hui Wang, Chin-Chen Chang, and Ming-Chu Li. Optimizing least-significant-bit substitution using cat swarm optimization strategy. *Information Sciences*, 192:98–108, 2012.

[16] Zhenfei Zhao, Hao Luo, Zhe-Ming Lu, and Jeng-Shyang Pan. Reversible data hiding based on multilevel histogram modification and sequential recovery. *AEU-International Journal of Electronics and Communications*, 65(10):814–826, 2011.

[17] Manoj Kumar Ramaiya, Naveen Hemrajani, and Anil Kishore Saxena. Security improvisation in image steganography using des. In *Advance Computing Conference (IACC), 2013 IEEE 3rd International*, pages 1094–1099. IEEE, 2013.

[18] Suresh N Mali, Pradeep M Patil, and Rajesh M Jalnekar. Robust and secured image-adaptive data hiding. *Digital Signal Processing*, 22(2):314–323, 2012.

[19] Rufeng Chu, Xinggang You, Xtangwei Kong, and Xiaohui Ba. A dct-based image steganographic method resisting statistical attacks. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 5, pages V–953. IEEE, 2004.

[20] Hui-Yu Huang and Shih-Hsu Chang. A 9/7 wavelet-based lossless data hiding. In *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2011 IEEE Symposium on*, pages 1–6. IEEE, 2011.

[21] Elham Ghasemi, Jamshid Shanbehzadeh, and Bahram ZahirAzami. A steganographic method based on integer wavelet transform and genetic algorithm. In *Communications and Signal Processing (ICCSP), 2011 International Conference on*, pages 42–45. IEEE, 2011.

[22] Phil Sallee. Model-based steganography. In *IWDW*, volume 2939, pages 154–167. Springer, 2003.

[23] Hirohisa Hioki. A data embedding method using bpcs principle with new complexity measures. In *Proc. Pacific Rim Workshop on Digital Steganography 2002 (STEG'02), Kitakyushu, Japan, July*, pages 30–47, 2002.

[24] Joachim J Eggers, Robert Baeuml, and Bernd Girod. Communications approach to image steganography. In *Security and Watermarking of Multimedia Contents*, pages 26–37, 2002.