Summarization of Typical JPEG Image Steganography and Steganalysis Methods

Lei Yu

College of Cryptographic Engineering, Engineering University of People's Armed Police, Xi'an 710086, China

ly1a2b3c@163.com

Keywords: steganography; steganalysis; JPEG image

Abstract: JPEG image is a very prevalent image format. With the development and application of JPEG image steganography techniques, JPEG image steganalysis research also develops. The development of steganography and steganalysis in confrontation. In this paper, the typical JPEG image steganography and steganalysis methods are outlined in brief. The development trend of JPEG image steganography and steganalysis research is also discussed to supply reference to the researchers.

1. Introduction

Steganography [1] is the art and science of “invisible” communication, which is to conceal the very existence of hidden messages. Images have many attributes, which make it suitable for steganography. Images can convey a large size of message. Because the non-stationarity of images, the image steganography is hard to attack. Especially, as the interchange of digital images is frequently used nowadays, image steganography becomes promising. Now, research in the field of JPEG steganography has become active as JPEG images are used popularly. Many steganographic techniques operating on JPEG images have been published and become publicly available. These steganographic techniques threatened information security, so JPEG images steganalysis research also develops.

Steganalysis is the art of detecting the presence of hidden messages, which is the counter problem to steganography. Although the presence of the embedded messages is often imperceptible to the human eye, it may nevertheless change the statistical properties of the cover image. Because of their invasive nature, steganographic systems often leave detectable traces within some characteristics. Certainly, the same is true of JPEG steganography. To attack JPEG steganography, two categories of steganalytic methods, includely the specific steganalysis and the blind steganalysis. A specific steganalysis method would give very good results when tested only on that embedding and might fail on all other steganographic algorithms. A blind steganalysis method might perform less accurately overall but still provided acceptable results on new embedding algorithms. Steganography algorithms wouldn't follow the Kerckhoff principle, so the blind steganalysis is more important to the specific steganalysis.

2. JPEG steganography

All JPEG steganographic techniques that embed messages in the image data can be broadly divided into three categories, includely JPEG input, side information and alternative domain. Each one of them can incorporate a different set of design elements [2].

2.1 JPEG input

JPEG input methods start with a JPEG `, extract the quantized DCT coefficients, modify them in order to embed the secret message, and then reassemble the stego JPEG file. The coefficients are usually determined using a selection rule and then a subset of them is modified using a predefined embedding opertaion[2]. For instance, JSteg, F5, OutGuess, Steghide, Model based (MB) steganography and so on.

Derek Upham’s JSteg [3] was the first publicly available steganographic system for JPEG images. Its embedding algorithm sequentially replaces the least-significant bit of DCT coefficients with the message’s data. The algorithm does not require a shared secret, as a result, anyone who knows the steganographic system can retrieve the message hidden by JSteg.

F5 [4] was developed form JSteg, F3 and F4. JPEG is the only image format that F5 works with. F5 takes two main actions to increase the security against steganalysis attacks: straddling and matrix coding. Straddling scatters the message as uniformly aspossible over the cover image to equalize the change density. With matrix embedding, F5 improves the embedding efficiency that is defined as the number of bits embedded per change of block DCT coefficient. Generally speaking, the smaller the embedding message size is, the larger the embedding efficiency of F5 is.

OutGuess [5] constructs a universal steganographic framework, which embeds hidden data using the redundancy of a cover image. For JPEG images, OutGuess preserves statistics of the block DCT coefficient histogram. Two measures are taken to reduce the change on the cover image introduced by data embedding. Before embedding, OutGuess identifies the redundant block DCT coefficients which have least effect on the cover image and will be modified if necessary during the data embedding. It also adjusts the untouched coefficients during the embedding procedure to preserve the original histogram of the block DCT coefficients after embedding.

Steghide [6] uses a graph-theoretic approach to steganography based on the idea of exchanging rather than overwriting pixels. It constructs a graph from the cover data and the secret message. Pixels that need to be modified are represented as vertices and possible partners of an exchange are connected by edges. An embedding is constructed by solving the combinatorial problem of calculating a maximum cardinality matching. The secret message is then embedded by exchanging those samples given by the matched edges. This embedding preserves first-order statistics. Additionally, the visual changes can be minimized by introducing edge weights.

MB [7] [8] embedding tries to make the embedded data correlated to the cover image. This is realized by splitting the cover image into two parts, modeling the parameter of the distribution of the second part given the first part, encoding the second part using the model and to-be-embedded message, and then combining the two parts to form the stego image. In embedding method MB, which operates on JPEG images, a Cauchy distribution is used to model the JPEG block DCT mode histogram. The embedding procedure keeps the lower precision version of the block DCT mode histogram unchanged.

2.2 Side information

Side information methods either require the input image to be in the raw uncompressed format and then embed the message while compressing the image by minimizing the combined distortion due to quantization and embedding or they manufacture the side information by repeated JPEG compression. Depending on the details of their embedding mechanism, these methods may or may not allow the use of matrix embedding. For instance, Perturbed Quantization (PQ) and Modified Matrix Encoding (MME).

In PQ [9], the sender hides data while processing the cover object with an information reducing operation that involves quantization, such as lossy compression, downsampling, or A/D conversion. The unquantized values of the processed cover object are considered as side information to confine the embedding changes to those unquantized elements whose values are close to the middle of quantization intervals. This choice of the selection channel calls for wet paper codes as they enable communication with nonshared selection channel.

MME [10] uses modified matrix encoding to choose the coefficients whose modifications introduce minimal embedding distortion. This method derives the expected value of the embedding distortion as a function of the message length and the probability distribution of the JPEG quantization errors of cover images.

2.3 Alternative domain

Alternative domain methods embed the message in a different domain robustly (e.g., in the spatial or wavelet domain) and then compress the image at the very end. On the one hand, the JPEG compression masks to a large extent the impact of embedding and the steganalyst can no longer inspect the direct impact of embedding changes. On the other hand, the compression introduces distortion and thus corrupts the message. Thus, the message needs to be embedded robustly so that the payload can be recovered without errors at the receiver. An example of this design element is YASS: Yet another steganographic scheme.

YASS [11] [12] uses a QIM-like mechanism to embed the message in selected bands of DCT coefficients of randomly positioned 8×8 blocks. After embedding, the image is compressesed and the stego image is advertised as JPEG. Robustness to JPEG compression is achieved by enlarging the payload using repeat-accummulate error correction codes before embedding to guarantee error-free extraction from the compressed image.

3. JPEG steganalysis

Steganalysis is the counter problem to steganography. It includes specific steganalysis and blind steganalysis. A specific steganalysis method would give very good results when tested only on that embedding and might fail on all other steganographic algorithms. A blind steganalysis method might perform less accurately overall but still provided acceptable results on new embedding algorithms.

3.1 JPEG specific steganalysis

For JSteg, Fridrich *et al* [14] presented RS attack which can detect it reliably. Westfeld [13] proposed a generic methodology to prepare higher order steganalystic methods form spatial domain for application in the transformed domain. He presented 72 new systematically designed methods that are derived form the spatial domain.

For F5, Fridrich *et al* [15] presented a steganalytic method. The key element fo the method is estimation of the cover image histogram from the stego image. The statistics of the original image were estimated by decompressing the JPEG image followed by cropping the four rows and four columns on the boundary, and then recompressing the cropped image to JPEG format using the original quantization table. The author claimed that the obtained image has statistical properties very much similar to that of the cover image. The number of relative changes introduced by F5 is determined using the least square fit by comparing the estimated histograms of selected DCT coefficients with those of the stegoimage.

For OutGuess, Fridrich *et al* [16] described new methodology for developing steganalytic methods for JPEG images. And they demonstrated the concepts by presenting a detection method for OutGuess. In the attack on OutGuess, they use the fact that the embedding mechanism in OutGuess was overwriting the LSBs. This means that embedding another message into the stego image will partially cancel out and will thus have a different effect on the stego image than on the cover image.

For YASS, Li *et al* [17] have present that the success of YASS is attributed to its innovation in embedding, i.e., hiding data in embedding host blocks whose locations are randomized. However, they find that the locations of the embedding host blocks are not randomized enough. Some locations in an image are possible to hold an entire embedding host block and some locations are definitely not. Additionally, YASS employs a Quantization Index Modulation (QIM) embedding strategy in order to enhance the robustness of the embedded data, which on the other hand introduces extra zero coefficients into the embedding host blocks during data hiding. Consequently, statistical features extracted from locations which are possible to hold embedding host blocks are different from those from locations which are impossible to hold embedding host blocks. The trace of YASS embedding is therefore exposed.

3.2 JPEG blind steganalysis

The blind steganalysis method for JPEG image can be broadly divided into three categories: (1) steganalysis method in the wavelet domain, (2) steganalysis method in the spatial domain and DCT domain, (3) steganalysis method in the DCT domain. The former two methods are effective to spatial image steganography algorithms and JPEG image steganography algorithms. The latter method is only effective to JPEG image steganography algorithms and the detection accuracy exceeds the former methods for JPEG image steganography algorithms.

(1) steganalysis method in the wavelet domain

Farid *et al.* [19] [20] proposed a universal steganalyzer based on image's higher-order statistics. Holotyak *et al.* [22] presented a universal statistical steganalysis of additive steganography using wavelet higher-order statistics. Its features are calculated from an estimation of the stego signal obtained from stego images in the wavelet domain. Xuan *et al.* [21] presented a universal steganalysis system. The statistical moments of characteristic functions of the image and their wavelet subbands are selected as features.

(2) steganalysis method in the DCT domain

Fridrich [23] has proposed a set of distinguishing features from the DCT domain and spatial domain. The statistics of the original image were estimated by decompressing the JPEG image followed by cropping the four rows and four columns on the boundary, and then recompressing the cropped image to JPEG format using the original quantization table. The author claimed that the obtained image has statistical properties very much similar to that of the cover image. Features for steganalysis were generated from the statistics of the JPEG image and its estimated version. Shi *et al* [24] modeled the differences between absolute values of neighboring DCT coefficients as a Markov process. The feature calculation started by forming the matrix JPEG 2-D of absolute values of DCT coefficients in the image. Four difference arrays were calculated along four directions: horizontal, vertical, diagonal, and minor diagonal. From these difference arrays, four transition probability matrices were constructed as features. Pevny *et al* [25] proposed an extended version of Fridrich's features which considered several different models for DCT coefficients and used the sample statistics of the models as features. Huang *et al* [26] presented an improved calibration-based universal JPEG steganalysis, where the microscopic and macroscopic calibrations were combined to calibrate the local and global distribution of the quantized block DCT coefficients of the test image.

(3) steganalysis method in the spatial domain and DCT domain

Lie *et al.* [27] proposed a feature classification technique, based on the analysis of two statistical properties in the spatial and DCT domains, to determine the existence of hidden messages in an image. Kodovsky *et al.* [18] argue that modern blind steganalysis tools recently developed for detection of JPEG and spatial domain steganography are capable of reliably detecting various settings of YASS even for small payloads and small images. They compare the detection rates with other steganography algorithms to put this intriguing steganographic algorithm in perspective.

4. Conclusion

The past few years have seen an increasing interest in using images as cover media for steganographic communication. There have been a multitude of public domain tools available for image based steganography. Given this fact, detection of covert communications that utilize images has become an important issue.

References

[1] Chandramouli R, Kharrazi M, Memon N. Image steganography and steganalysis concepts and practice. Proceedings of 2nd Intemation Workshop on Digital Watermarking. Seoul, South Korea: Springer, 2003, 2939:35~49.

[2] Kodovsky J, Fridrich J. Influence of embedding strategies on security of steganographic methods in the JPEG domain. Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X. San Jose, CA, 2008, 6819:1~13.

[3] UphamD. JSteg, http://www.funet.fi/pub/crypt/steganography/jpeg-JSteg-v4.diff.gz

[4] Westfeld A. F5 – a steganographic algorithm. Proceedings of the 4th International Workshop on Information Hiding. 2001, 2137:289~302.

[5] Provos N. Defending against statistical steganalysis. Proceedings of the 10th USENIX Security Symposium. Washington, D.C., USA: IEEE, 2001, 323~335.

[6] Hetzl S, Mutzel P. A graph–theoretic approach to steganography. Proceedings of Communications and Multimedia Security 9th IFIP TC-6 TC-11 International Conference (CMS 2005). Salzburg, Austria:Springer, 2005, 3677:119~128.

[7] Sallee P. Model-based steganography. Proceedings of Digital Watermarking 2nd International Workshop (IWDW 2003), Seoul, Korea:Springer, 2004, 2939:154~167.

[8] Sallee P. Model-based methods for steganography and steganalysis. International Journal of Image and Graphics. 2005, 5(1):167~189.

[9] Fridrich J, Goljan M, Soukal D. Perturbed quantization steganography. ACM Multimedia and Security Journal, 2005, 11(2):98~107.

[10] Kim Y, Duric Z, Richards D. Modified matrix encoding technique for minimal distortion steganography. Proceedings of 8th International Workshop on Information Hiding. New York: Springer, 2006, 4437:314~327.

[11] Solanki K, Sarkar A, Manjunath B S. YASS: yet another steganographic scheme that resists blind steganalysis. Proceedings of 9th International Workshop on Information Hiding. Saint Malo, Brittany France: Springer, 2007, 4567:16~31.

[12] Sarkar A, Solanki K, Manjunath B S. Further study on YASS: steganography based on randomized embedding to resist blind steganalysis. Proceedings of SPIE-Security, Steganography, and Watermarking of Multimedia Contents X. San Jose, CA, USA: SPIE, 2008.

[13] Westfeld A. Generic adoption of spatial steganalysis to transformed domain. Proceedings of 10th International Workshop on Information Hiding. USA: Springer, 2008: 161~177.

[14] Fridrich J, Goljan M, Du R. Detecting LSB steganography in color and gray-scale Images. IEEE Multimedia, 2001, 8(4), 22~28.

[15] Fridrich J, Goljan M, Hogea D. Steganalysis of JPEG image: breaking the F5 algorithm. Proceedings of 5th International Workshop on Information Hiding, Noordwijkerhout, Netherlands:Springer, 2002, 310~323.

[16] Fridrich J, Goljan M, Du R. Attacking the OutGuess. Proceedings of the ACM Workshop on Multimedia and Security, Juan-les-Pins, France:ACM, 2002, 3~6.

[17] Li B, Shi Y Q, Huang J W. Steganalysis of YASS. Proceedings of the 10th ACM Multimedia & Security Workshop. Oxford: ACM, 2008:139~148.

[18] Kodovsky J, Pevny T, Fridrich J. Modern Steganalysis Can Detect YASS. Proceeding of SPIE Electronic Imaging, Media Forensics and Security Ⅻ. San Jose: SPIE, 2010:1~11.

[19] Farid H. Detecting hidden messages using higher-order statistical models. Proceedings of the 5th Intl. Conf. on Image Processing. New York, USA, 2002, 2:905~908.

[20] Farid H, Siwei L. Detecting hidden message using higher-order statistics and support vector machine. Proceeding of 5th Information Hiding Workshop. Noordwijkerhout, Netherlands:Springer , 2002, 2578: 131~142.

[21] Xuan G R, Shi Y Q, Gao J J, et al. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. Proceeding of 7th Information Hiding Workshop. Barcelona, Spain:Springer , 2005, 3727:262~277.

[22] Holotyak T, Fridrich J, Voloshynovskiy S, Blind Statistical Steganalysis of Additive Steganography Using Wavelet Higher Order Statistics, *9th IFIP TC-6 TC-11 Conference on Communications and Multimedia Security*, LNCS vol. 3677, Springer-Verlag, Berlin, pp. 273–274, 2005.

[23] Fridrich J. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. Lecture Notes in Computer Science, 2005, 3200: 67~81.

[24] Shi Y Q, Chen C, Chen W. A Markov process based approach to effective attacking JPEG steganography. Proceedings of Information Hiding Workshop 2006. Heidelberg:Springer, 2006: 249~264.

[25] Pevny T, Fridrich J. Merging Markov and DCT features for multi-class JPEG steganalysis. Proceedings of SPIE Electronic Imaging, Security, Steganography, and Watemarking of Multimedia Contents IX. San Jose: SPIE, 2007:3~4.

[26] Huang F J, Huang J W. Calibration based universal JPEG steganalysis, Science in China Series F: Information Sciences, 2009, 52(2): 260~268.

[27] Lie W, Lin G. A feature-based classification technique for blind image steganalysis. IEEE Transactions on Multimedia. 2005, 7(6):1007~1020.