Analysis and Categorical Application of LSB Steganalysis Techniques

Joshua Cazalas, Todd Andel and Jeffrey McDonald
University of South Alabama, Mobile, USA
jdc703@jagmail.southalabama.edu
tandel@southalabama.edu
jtmcdonald@southalabama.edu

Abstract. Many tools and methods for steganalysis are prevalent in the research field. While no technique is 100% effective, combining multiple techniques is common practice. Techniques reliant on the same basis are often found to be less computationally efficient when used in combination as opposed to the combined use of techniques in separate categories. It is the goal of this paper to present many of the foundational techniques, explore their strengths and weakness, categorize the techniques, and present further theories on the combinational effectiveness of techniques within and outside of those categories.

Keywords: steganalysis, LSB, steganography

1. Introduction

Steganography, the act of hiding communications in their entirety is often found in varying forms of digital media. One of the primary media targets for this communication and information hiding is digital images, although the techniques apply to music, file header information, film, and many other forms of digital media. One of the most popular techniques in steganography is the use of least significant bit (LSB) hiding, which is the act of changing the least significant bits of digital media to hide communications.

Many techniques exist for analyzing media believed to contain steganographic communications. The study of discovering the communications hidden using steganography is known as steganalysis. Most of these techniques rely heavily on statistics, matching, optimization, and at least some knowledge that information or communications are being hidden in a given source. It should be apparent that if steganography is actually successful in hiding the fact that communications are taking place at all, then those performing steganalysis have no means of knowing their techniques are needed.

Techniques for steganalysis include sample pair analysis, discrete cosine transform (DCT) statistics, joint probability distributions, and matching. The commonality found between all these techniques are some type of statistical analysis to determine the probability that the digital information distributions would naturally occur. Some protocol combinations produce higher rates of success when performing steganalysis on different types of digital media; however, some combinations may actually lead to incorrect analysis.

The purpose of this study is to compare and contrast the success, benefits, and ease of implementation or understanding of both individual steganalysis techniques and combinations of those techniques. The study focuses on the maximizing of success rates in image steganalysis by optimizing the techniques or changing the techniques used based on the statistical data provided by a set of standardized techniques.

When steganography techniques are used in combination, much like cryptography techniques, the overall level of obscurity added to the information being hidden is often increased. It is the belief of these authors that uncommon combinations of steganalysis techniques could optimize techniques for analysis on specific media types.

The remainder of the paper is as follows. Section II is a presentation and review of the steganalysis techniques commonly used. Section III contributes a taxonomy by which steganalysis techniques can be classified, and it defends those classifications. Section IV is a synopsis of other related research, and it is followed by the conclusion and result.

2. Steganalysis techniques

A large variety of techniques for determining whether or not a piece of digital media contains a steganographic message exist. However, the most commonly encountered techniques analyze statistical data based on the expected distribution and actual distribution of the media’s LSB streams [5, 7].

LSB steganography has become a very common means of hidden message passing through open channels. It is particularly prevalent in digital images because of the nature of image bit streams. The message passer simply alters the LSB stream of an image in some predetermined manner that can be extracted by the intended recipient and decoded. This change is rarely noticeable to the human eye. Depending on the size and type of image or media, the change is often untraceable and undetectable through standard computational analysis unless it is already known that a message has been hidden within the image or media.

The techniques under analysis for our study include sample pair analysis (Dumitrescu, 2003), regular and singular groups (Fridrich, 2001), difference image histogram (Tao, 2003), joint density (Pevny, 2007), Markov processing (Shi 2006), category attack (Lee, 2006), and joint probability (Liu, 2011). The common thread in all these approaches is that they generally require some prior statistical data about the media type being analyzed.

2.1 Sample pair analysis

The technique of sample pair analysis (SPA) was developed in 2002-2003 by Sorina Dumitrescu (2003) and aimed to show that “the length of hidden message embedded in the least significant bits of signal samples can be estimated with relatively high precision.”

What this meant for practical steganalytic techniques was that if one could determine that a message had been passed using LSB steganography then its length could be determined. This type of information is able to provide the statistical data needed by other steganalysis techniques to be performed with higher rates of precision. The approach was based on using statistical measures of specific sample pairs that are highly sensitive to LSB alterations or embedding. The algorithm was trained on existing media until it was considered robust and the estimation errors could be determined.

Steganographers were able to attack this analysis technique and new techniques were generated to avoid detection within a rather short time frame. Algorithms that determined more chaotic distribution of their messages (Xiangyang, 2006) that could be shared with the recipient were sufficient to deal with the SPA method and have it return falsely low estimations on possible hidden message lengths.

2.2 Difference image histograms

The technique of difference image histogram (DIH) analysis was one accepted as a thorough approach for analyzing raw lossless image types for a number of years. DIH is still used with some frequency in preliminary analysis by steganographers. IT is used to produce an image histogram as a graphical representation of the tonal distribution in a digital image.

Steganalytic techniques can take advantage of this graphical representation in many ways. In 2003, Zhang (Tao 2003) proposed a technique in which “translation coefficients between difference image histograms are defined as a measure of the weak correlation between the least significant bit plane and the remaining bit planes.” This correlation can be used to construct classifiers that discriminate between stego-images and carrier-images. What was discovered in his study was that the algorithm he developed could not only detect the existence of LSB steganography in digital images with extreme reliability, but was also able to predict the number of steganographic messages within a digital image with extreme precision.

Using DIH, the difference image is defined as (Kekre, 2011):

$$D(i, j) = X(i+1, j) - X(i, j)$$

(Equation 1)

where X (i, j) denotes the value of the image X at the position (i, j). Zhang and Ping found that there exist differences between the DIH for normal images and the images obtained after flipping operation on the LSB plane (Tao, 2003).

The computation speeds for DIH steganalysis, while still intensive, had much lower benchmarks than SPA and other similar differencing techniques. However, his technique, like any security or de-obfuscation technique, is not universally secure and a study performed in 2011 (Kekre, 2011) presented an algorithm specifically designed to thwart this steganalysis technique.

The technique used to subvert this analysis technique is based on normal statistical distribution of bits in the LSB plane. Any normal image's LSB plane should consist, on average, of an equal number of ones and zeros (Kekre, 2008). The DIH algorithm works by flipping all the bits in the LSB plane to zeros and then doing a difference comparison with the original image. If the steganographic message is tailored to embed more zeros than ones it is possible to create a DIH that appears to denote a natural image. This conclusion allowed Kekre to develop the following algorithm to escape standard DIH steganalysis. (Kekre, 2008):

Input: A cover bitmap (BMP) image to convert into a stego image and a BMP image to hide in the cover image. The size of the image to be hidden should be approximately one eighth of the cover image's size.

Output: A stego BMP image.

Step 1: Convert the cover image into gray scale.

Step 2: Convert the BMP image to be hidden into gray scale.

Step 3: Change the LSB of each pixel of cover image to zero.

Step 4: Sequentially replace the second least significant bit of each pixel of cover image by one bit of the hidden image. That is eight pixels of the cover image will contain information about one pixel of hidden image.

2.3 Regular and singular groups

By inspecting the differences in the number of regular and singular groups (RS) for the LSB and the "shifted LSB plane," it is possible to detect even very short messages in 24-bit color , 8-bit color, or grayscale images (Fridrich 2001).

Fridrich defines three groups: Regular, Singular, and Unusable; based on a discrimination function *f* (2001):

$$f(x_1, x_2, \ldots, x_n) = \sum^{n-1} |x_{i+1} - x_i|$$

(Equation 2)

and a flipping operation *F*.

$$F_{-1}(x) = F_1(x+1) - 1 \ \mathit{for\ all}\ x$$

(Equation 3)

Regular groups: G □ R □ f(F(G)) > f(G) Singular groups: G □ S □ f(F(G)) < f(G) Unusable groups: G □ U □ f(F(G)) = f(G)

The basic premise is that a mask is applied to the Regular Group and again to the Singular Group. The difference of these two groups should be zero even if the entire plane is shifted by one unless the LSB plane has been modified or had additional data embedded. In theory this method works on the same principles as DIH.

2.4 Markov processing

The Markov process is a well-established statistical and probabilistic theory in which one can predict the future of a process based solely on its present state. The properties of this particular theory have proven useful particularly in digital image steganalysis. Shi (2006) presented a process for modeling JPEG image vectors using the Markov statistical process and proceeds to use those models as a measurement for determining if an image has had a steganographic message embedded within its LSB stream or in other bit planes.

The original proposal for the Markov process for steganalysis involved the modeling of 2D arrays of JPEG images formed by the magnitudes of the DCT coefficients. These were used as difference arrays to which the Markov process could be applied to produce second order statistical values. The original Markov process also involved a trained algorithm threshold to reduce computation needed to compute the dimensionality matrices involved in the Markov process.

Results of the Markov processing for steganalysis were extremely promising against some of the more prominent JPEG steganography techniques: OutGuess, F5, and MB1 with open source algorithms available [12, 13, 14]. It was noted that certain hidden images still eluded detection by these algorithms and more research was needed to determine the similarities between those images.

2.5 Joint density

Joint Density steganalysis is an expansion of the Markov process techniques based on the properties of dependency between compressed DCT coefficients and their neighboring values in the Markov process (Pevny, 2007). It was determined that hiding information, even if it does not alter the zero values analyzed by the Markov process, will in fact modify the neighboring joint density of JPEG images (Liu, 2011). The combination of the standard Markov process and the analysis of the joint neighbor densities is determined to "simultaneously achieve faster detection time, and higher detection performance for JPEG image steganography" than standard Markov process steganalysis (Arun).

2.6 Category attack

When a Markov-based process is used, it takes advantage of the second order statistics of DCT coefficients to perform steganalysis. The category attack presents a similar process which uses the first order statistics of these values (Lee, 2006). While perhaps not as powerful as the Markov processing approach, it is an improvement on the Zhang and Ping histogram method for the majority of steganalysis categories.

The category attack simply exploits the histogram of DCT coefficients, but is more powerful to detect the randomized Jsteg (Upham) embedding as well as the randomized Jphide (Latham) embedding than the DIH steganalysis technique. "The detection power of both proposed methods were compared to the blind classifier by Fridrich that uses 23 DCT features" to reach this conclusion (Lee, 2006).

3. Categorical analysis

Steganography, much like cryptography and security in general, is an ever adapting field of study. As steganography changes the techniques used in steganalysis will likewise change in their efficacy to detect hidden messages or to recover hidden messages if that is their expected result. It should be apparent that while many of the previously explored techniques utilized some form of similar axioms to begin their approach, statistical data over time and the evolution of steganographic techniques altered their conclusions and final algorithms drastically. This section categorizes those algorithms based on ease of implementation, efficacy in the detection of steganographic messages, and joint usage for the amelioration of existing techniques. The categories developed here are primarily centered on the weaknesses found in a given steganalysis technique. The reasoning behind this approach is to improve the joint use of multiple techniques. If two techniques suffer from the same weakness it stands to reason that the combination of the two techniques will further suffer from that weakness and continue to be vulnerable to the same or similar failures.

We propose the following categories to classify steganalysis techniques broadly based on the previously identified properties. (1)

Table 1: Categories

Category	Attributes	Algorithms
Zero Distribution	Binary Distribution Statistics Output Hidden Length	Sample Pair Analysis Regular and Singular Groups Difference Image Historgrams
Vector Plane Processing	Statistical Model Representation Image Type Specific Geometric/Trigonometric	Markov Processing Category Attack
Abstract Parellel Correlations	Abstracted Statistics Correlated Statistics Preprocessed Data Detect Specified False Negatives	Joint Density Joint Probablity

3.1 Zero distribution

The first category of steganalysis techniques being proposed is the Zero Distribution Steganalysis category. The algorithms in this category use statistical data about the distribution of ones and zeros in an LSB plane or another bit stream that is reliant on normal images containing an even distribution of ones and zeros.

Examples of steganalysis techniques that fall into this category include SPA, regular and singular groups analysis, and DIH analysis. All of these algorithms by default make the assumption that steganographic messages embedded in digital images alters the distribution of ones and zeros within the LSB plane or the bit stream into which the data was embedded.

For these three steganalysis techniques in this category a comparison and equivalence statement of their proofs has been created (Xiangyang, 2006). The extracted hypothesis for DIH is

$$a_{2m,2m+1}g_{2m} \approx a_{2m+2,2m+1}g_{2m+2}$$

(Equation 4)

which is found to be equivalent to the extracted hypothesis for SPA.

$$E\{||X_{2m+1}||\} = E\{||Y_{2m+1}||\}$$

(Equation 5)

Xiangyang proves the equivalence of the two hypothesis in these two methods by transforming the hypothesis in DIH in the following manner (Xiangyang, 2006):

$$a_{2m,2m+1}g_{2m} = \frac{\|B_{2m+1}\|}{\|G_{2m}\|}\|G_{2m}\| = \|B_{2m+1}\|$$

$$a_{2m+2,2m+1}g_{2m+2} = \frac{\|A_{2m+1}\|}{\|G_{2m+2}\|}\|C_{2m+2}\| = \|A_{2m+1}\|$$

Thus the hypthesis (4) of DIH method can be converted into

$$\|A_{2m+1}\| = \|B_{2m+1}\|$$

(Equation 6)

Xiangyang provides similar proofs to show equivalence for DIH and RS as well as for RS and SPA in the same publication.

This equivalency provides a basis for the theory that generalized steganographic techniques which cause failures in one of these algorithms should cause the same failures or similar failures in the other two algorithms. This may not always hold, and there may be failures in one algorithm that are not caused by the basis or hypothesis on which the steganalysis technique is based but rather on the exact implementation of the algorithm used to approach that technique.

Another commonality between the techniques included in this category is the output data types. These techniques and the algorithms that implement them in general output an expected length of the hidden messages rather than a presence or lack of steganographic information. These techniques could benefit from preprocessing that determines whether or not a message is to be expected in a few regularly occurring situations such as the avoidance algorithm discussed in Section II Part B of the paper. Other tweaks could be made to the algorithms themselves to deal with specific counters but the generalized attack method would still be successful against the steganalysis.

3.2 Vector plane processing

Vector Plane Processing is the second proposed classification category, which involves the transformation of the analyzed data type to be represented fully by statistical models. Steganalysis techniques that fall into this category include techniques that use first and second order statistical data of trigonometric transformations like the DCT. Markov processing and the category attack are two examples of techniques within this category of steganalysis.

These analysis techniques are often tailored toward a specific digital image type and focus on using models or representations of data that correlate to the LSB plane or some other bit stream; however, are not direct interpretations of the data found in those bit streams.

These techniques are often reliant on algorithms in the statistical process being computable in reasonable polynomial time, and in Markov processing and the category attack, on DCT features and properties. Techniques that use other transforms may fall into this category as well but would more than likely still be reliant on the planar geometric and trigonometric properties of that representation.

Lee was able to show a significant improvement in the results of the category attack when he applied many of the features used in the Markov processing (Lee, 2007); however, the images that remained outside the domain of recognition had intrinsic properties making them invulnerable to this type of analysis. That is to say, that a steganographic product of techniques in this category may prove more beneficial than one in previous categories in the actual recovery of images but will likely not see an improvement in the actual detection rate of hidden messages for a given data type.

Adaptations of steganalysis techniques within this category can often be targets for analyzing non common data types if they can be represented statistically and modeled in a corollary fashion to the existing prediction models.

3.3 Abstract parallel correlations

It is believed that there are techniques which fall into a category of Abstract Parallel Correlations. Steganalysis techniques in this category would bear resemblance to the discussed joint density analysis process.

While joint density steganalysis is an adaptation of a technique from the Vector Plane Processing category, in this case Markov processing, it uses correlated data that is an abstraction of the data provided by the Markov process. This approach is an example of related data from using one process having direct correlation to the problem being solved or attempted.

Techniques in this category either use preprocessing from another technique to get abstract data for their input streams or use correlated mathematic or statistical theories to make educated guesses about the expected results. These types of techniques are particularly useful in identifying steganographic messages in images or digital media that has been tailored specifically to avoid known detection algorithms. Oftentimes this involves tailoring the message or data to flag a false negative but does not change the associated statistical properties of that embedding.

Techniques in this category may not provide results as accurate as those in the other two categories; however, they are generally more independent and capable of recovering hidden messages that have been tailored specifically to avoid detection by the techniques in the other categories. These can provide statistical results that show another algorithm may have been manipulated to provide a false negative or false positive.

3.4 Categorical implications

The hypothesis about the proposed categorical separation is that the product or hybrid adaptation of compatible techniques from the separate categories would have a more positive effect on steganalysis results than the product or hybrid combination of techniques within the same category.

In some cases the algorithm implementing the steganalysis technique was the limiting factor. The types of tailored data that would negatively affect the results of the steganalysis on one technique in a category would have the same negative impact on all or most techniques in that category. This would imply that creating a product of steganalysis techniques within the same category could become computationally intensive with minimal increase in the efficacy of the techniques, but creating a hybrid or product of steganalysis techniques from different categories is much more likely to be beneficial to the overall result.

The Vector Plane Processing techniques are generally a good place to start if an algorithm exists in the category for the data type being analyzed. Techniques from the Zero Distribution category benefit greatly from the preprocessing done by the techniques in the Vector Plane Processing category. It is often beneficial to only run algorithms from the Zero Distribution category if it is already believed a hidden message exists, or in the case that the data type being analyzed is greatly dependent on statistically random distributions. When performing an analysis with a technique in the Abstract Parallel Correlations category most of the preprocessing should have been completed by the other techniques and there should be some indication that data is “too perfect” or using avoidance mechanisms to hide from steganalysis techniques in the other categories.

4. Related research

Other research exists to categorize steganalysis techniques and improve upon existing techniques using statistical models. One such example is a project at Oxford University under the supervision of Andrew Ker in which machine learning is being used to combine steganalysis techniques based on the results of embedded features recognized by various statistical methods.

“We can analyze (or simulate) the effect of embedding on features to a high level of detail. It turns out that some features are "equivalent" in that the same thing happens to them under embedding, in which case it is more efficient to combine them...The combined features can be benchmarked against the standard features using machine learning tools.” (Visualization of steganalytic features)

Research continues to expand in the creation of new steganalysis techniques however it is not uncommon for researchers to attempt to improve existing techniques using only data already produced by that technique without introducing the data provided by external steganalysis techniques.

5. Conclusion

The purpose of this study was to compare and contrast the success, benefits, and ease of implementation or understanding of both individual steganalysis techniques and combinations of those techniques. The results of this analysis show that some techniques are very closely related to other techniques in their hypothesis and basis which implies that their subversion can be done similarly or simultaneously. These subversion techniques often consist of manual tailoring or automated approximations. The steganalysis techniques which are the targets of these subversions were then categorized. It is believed that the product of steganalysis techniques from separate categories could detect the automated forms of subversion aimed at techniques from a given category whereas the product of steganalysis techniques from the same category may prove to be computationally intensive with minimal improvements in their results.

Future research that is planned at this time includes the analysis of various steganalysis technique products to verify the hypothesized results as well as adaptations of the classification model proposed. It is likely that more categories should exist and it is also apparent that there are many techniques that have not been given a classification in this survey and analysis.

References

Arun, R., S. Nithin Ravi, and K. Thiruppathi. "Neighboring Joint Density and Markov Process Based Approach for JPEG Steganalysis."

Dumitrescu, S.; Xiaolin Wu. 2005 "A new framework of LSB steganalysis of digital media," Signal Processing, IEEE Transactions on , vol.53, no.10, pp. 3936- 3947, Oct. 2005 DOI=10.1109/TSP.2005.855078 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1511008&isnumber=32356

Dumitrescu, Sorina; Wu, Xiaolin; Wang, Zhe. 2003 "Detection of LSB Steganography via Sample Pair Analysis" Information Hiding, 355-372 Springer Berlin / Heidelberg DOI=10.1007/3-540-36415-3_23 http://dx.doi.org/10.1007/3-540-36415-3_23

J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," Proc. ACM Workshop on Multimedia and Security, pp. 27–30, 2001.

H. B. Kekre, A. A. Athawale, and S. V. Maheshwari. 2011. Escaping difference image histogram steganalysis. In Proceedings of the International Conference & Workshop on Emerging Trends in Technology (ICWET '11). ACM, New York, NY, USA, 463-466. DOI=10.1145/1980022.1980121 http://doi.acm.org/10.1145/1980022.1980121

H. B. Kekre, Athawale A. Archana, "Information hiding using LSB technique with increased capacity", in International Journal of Cryptography and Security, vol.1, No.2, October 2008.

Allan Latham, JPHide, http://linux01.gwdg.de/~alatham/stego.html

Lee, Kwangsoo; Westfeld, Andreas; Lee, Sangjin 2006 "Category Attack for LSB Steganalysis of JPEG Images" Digital Watermarking, 35-48 Springer Berlin / Heidelberg Doi=10.1007/11922841_4 http://dx.doi.org/10.1007/11922841_4

Lee, Westfield Category attack for LSB steganalysis of JPEG images 2006

Lee, A. Westfeld. Generalized category attack - improving histogram-based attack on JPEG LSB embedding. In Information Hiding, 9th International Workshop, Saint Malo, France, June 11-13, 2007, Lecture Notes in Computer Science. Springer-Verlag, 2007.

Qingzhong Liu, Andrew H Sung, Mengyu Qiao, (2011) " Neighboring Joint Density-Based JPEG Steganalysis", ACM Transactions on Intelligent Systems and Technology. volume 2, No 2, Article 16.

Y. Q. Shi, C. Chen, and W. Chen. A markov process based approach to effective attacking jpeg steganography. In Information Hiding, pages 249–264, 2006

T. Pevny and J. Fridrich, (2007) "Merging Markov and DCT features for multi-class JPEG steganalysis", Proceedings of SPIE Electronic Imaging, Security, Steganography and Watermarking of Multimedia Contents IX, volume 6505, pages 650503-1 to 650503-13

Tao, Zhang; Xijian, Ping 2003 "Reliable detection of LSB steganography based on the difference image histogram" IEEE Xplore Conference Publication, Doi= 10.1109/ICASSP.2003.1199532

Derek Upham, JSteg, http://zooid.org/~paul/crypto/jsteg/

"Visualization of steganalytic features" https://www.cs.ox.ac.uk/teaching/studentprojects/238.html

Huaiqing Wang and Shuozhong Wang. 2004. Cyber warfare: steganography vs. steganalysis.Commun. ACM 47, 10 (October 2004), 76-82. DOI=10.1145/1022594.1022597 http://doi.acm.org/10.1145/1022594.1022597

Xiangyang, Luo; Chunfang, Yang; Fenlin, Liu 2006 "Equivalence Analysis Among DIH, SPA, and RS Steganalysis Methods" Communications and Multimedia Security Lecture Notes in Computer Science Volume 4237, 161-172 Springer Berlin / Heidelberg Doi=10.1007/11909033_15

Xiangyang, Luo; Zongyun, Hu; Can, Yang; Shanqing, Gao 2006 "A secure LSB steganography system defeating sample pair analysis based on chaos system and dynamic compensation" IEEE Xplore Conference Publication, Phoenix Park, Doi=10.1109/ICACT.2006.206144

T. Zhang, X. Ping. "A new approach to reliable detection of LSB steganography in natural images", Signal Processing, Elsevier, Volume 83 ,Issue 10, 2003, 2085-2093

http://www.outguess.org/

http://wwwrn.inf.tudresden.de/~westfeld/f5.html

http://redwood.ucdavis.edu/phil/papers/iwdw03.htm