IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol-6, Special Issue-6, July 2018 E-ISSN: 2347-2693

Steganalysis -Iterative Rule Learning to Discover Patterns

Megala G[1*], Maya Mohan[2]

[1]Computer Science and Engineering, N.S.S College of Engineering, Palakkad, Kerala
[2]Computer Science and Engineering, N.S.S College of Engineering, Palakkad, Kerala

[*]*Corresponding Author: megalaguruvayurkutty1@gmail.com*

Available online at: www.ijcseonline.org

***Abstract*—** In these years, everything is leading to new development and digitization. With these developments in technology, main challenge which exists is the threat of security. Steganography method usually embeds the sensitive messages in visually innocent cover images. The target of steganalysis is to determine the stego images from that of empty images. Every method depending on its hiding capacity of secret data in images place a unique markings or signature in stego images. To find this kind of markings in the images leads us to encorporate a classifier to be made for the purpose of finding the stego images which are usually the outcome of such steganography algorithm. In this, approach involves an evolutionary fuzzy rules to take out the markings of stego images in contrast to those empty images. Thus by using knowledge discovered, appropriate models for steganalysis can be involved and stego images can be found out and evolutionary algorithm can be optimized well. Thus the particular signature of steganographic method can be taken out well and also the kind of method used to produce stego image can be predicted.

***Keywords*—** Steganalysis, Fuzzy rules, Evolutionary Genetic Algorithm, Iterative Rule Learning.

I. Introduction

Steganalysis is a kind of analysis done inorder to find or extract the hidden message from the carrier cover images. It is a collection of techniques: which can determine the exixtence of stego content in cover image by visual or statistical[1]. Steganalysis can also be categorized to be passive or active. Passive steganalysis as it name specifies is to only find out the existence or absence of confidential message whereas active steganalysis[12][13] try to estimate and determine the text length, secret key, text bits, etc and various other information. Recent steganalysis techniques mainly focus the classifier design build on the training set of carrier images and stego images generated from various different embedding algorithms. When an image undergoes some kind of implanting process; the inherent or statistical features get changed. Thus classification can be done based on those features. The principal task of steganalysis lies in the removal of sensitive features and model of good classifier.

This method uses an evolutionary process to take out the markings of stego images from that of empty images using fuzzy rules. From these obtained knowledge, appropriate models for steganalysis can be involved. An Evolutionary algorithm (EA)[2] are collection of evolutionary computations. Evolutionary algorithm which is based on Iterative Rule Learning(IRL) approach is used for rule generation. The rules are generated incrementally so that the evolutionary algorithm can able to optimize one fuzzy rule at a time.

The formed fuzzy rules are used as the markings of stego images or, in other words, pattern of the underlying method used for embedding. The concept of fuzzy rules are used, ie fuzzy logic has been described as an intelligent alternative for modeling and control of complex systems in a practical, robust way.

The sections of the paper is organized as follows. Section II presents the related works being done, Section III presents the system architecture, including generating stego images, feature extraction, evolutionary fuzzy system, and evolutionary fuzzy algorithm and we conclude the paper with discussions in Section IV.

II. RELATED WORK

Steganalysis was initiated inorder to determine the undisclosed messages embedded in image or video[3]. In steganalysis the analysis can be divided into: visual based analysis and statistical based analysis. Visual analysis mainly based on finding confidential message with naked eye or with the help of computer which is done by analyzing the bit planes for any undetectable change in the appearance of undisclosed message. Statistical analysis deals this with finding any change in statistical properties of stego object which is usually triggered by steganographic algorithm. The effectiveness of the steganalysis system is mainly based on method used in feature extraction and also selection of the classifier for analysis.

To reveal the presence of confidential message in an image different features of images are used by numerous steganalysis method. Generally, feature selection can be categorized into two category: filtering and wrapper methods. Filtering methods do not include learning and from learning classifiers they select features independently. The drawback of filtering methods is that they doesn't consider the possible interaction of features among them and consider single features in isolation. Although, the combination of these features may have a consolidate effect which doesn't follow from the separate interpretation of features in the group. Whenever, there is a limitation on the number of features to be selected, efficient features may not be included. The wrapper method works based on feature groups regarding the found out classification errors and then model the final classifiers. Each steganography method involves a particular way to inject secret data in the images. Hence, they puts a unique markings on the stego images. To identify these unique markings in the images uses an appropriate classifier to be build mainly to find those stego images.

III . SYSTEM ARCHITECTURE

The method consists of two stages: Training phase and the Testing phase.

A. Training Phase

In the first stage, we examine an image database to explore the pattern or signature of stego images. Stego images are created by encrypting the images using the three steganography method. Mainly PQ, MB, YASS and these method cause very less detectable artifacts in images. Model based Steganography method is general, and can be applied to virtually any type of media. It provides answer for some fundamental questions which was not be able to fully addressed by previous steganographic method, such as how large a message can be hidden without detection by using statistical methods, and how to achieve this maximum capacity. YASS, reffers to Yet Another Steganographic Scheme chose random locations in images to conceal text so as to nullify the process such as, cropping usually adopted by blind steganalysis schemes. This reduces the self calibration process in the images while processing the images. Perturbed Quantization, is another method in which text is hide when processing the carrier image with operations quantization, lossy compression.

Fig 1 shows that images are trained from the image database which consist of both clean and stego images. Patterns which are the useful features of stego images and their respective values. This pattern are generated as set of fuzzy if-then rules that implies the likeness between stego images. Mainly histogram of spatial representation and discrete wavelet representation features are considered, because of its statistical artifacts on the images. Then these features are normalized into the unit interval[0,1] to use the identical membership function in the fuzzy rule generation. Evolutionary fuzzy systems[9] are used to find the signature of the stego images because of their unambiguous and intelligible knowledge. Signature corresponds to the patterns of the stego and clean images are represented as a stego or clean fuzzy rules. These fuzzy if-then rules are generated using membership functions. And by using IRL, optimizes fuzzy rule in each iteration of the algorithm and iteratively learns rule as progress.[10]. Finally fuzzy rules are obtained as the signature of images as clean and stego rules.

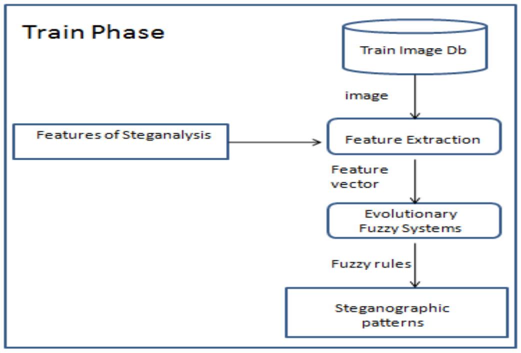

Figure 1. Training Phase

B. Testing Phase

In testing phase Fig 2 on giving a test image, corresponding features are extracted and a feature vector is produced. Then using the feature vector, checks the currently obtained pattern with already obtained pattern rules obtained during the learning process. During the learning process already a set of fuzzy rules for clean and stego are obtained. So this feature vector is matched with these all patterns to find the best match. Only the particular model which matches the pattern best is used to employed.

Instead of checking a test image to all models, we can pair its feature vector with all the signature and after deciding the most appropriate one, the corresponding model can be engaged. In other words, after exploring the patterns of the employed steganography method from the stego image, suitable model is used to evaluate the type of the method used in the process. Thus reduces the evaluation on all the models. The advantage of using genetic algorithm is that it can optimize values and could bring better learning. This could be improved further ie, after detecting the type of steganography method, the message bits, message length and embedded key can be estimated as a part of active steganalysis. Since specific steganalysis method is used to detect the employed methods, it can utilize the full knowledge of steganographic algorithms and can be extended to active steganalysis.

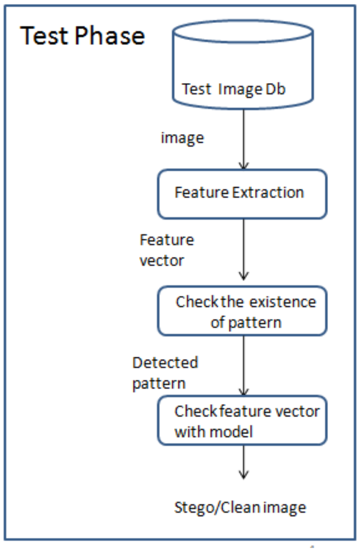

Figure 2. Testing Phase

C. *Generating Stego Images*

The stego images for training the image database was done by obtaining JPEG images from Washington University image database and then converted to grayscale images. Then to form stego image databases, MB,[7] PQ,[8] and YASS[6] steganography methods are employed. Model based Steganography method is general, and can be applied to virtually any type of media. It provides answer for some fundamental questions which was not be able to fully addressed by previous steganographic method, such as how large a message can be hidden without risking detection by certain statistical methods, and how to achieve maximum capacity. Yet Another Steganographic Scheme(YASS), is a scheme which performs with the aim is to negotiate the process of cropping in random locations of images which is usually done by blind steganalysis schemes. In Perturbed Quantization, is another method in which text is embedded when processing the carrier image with operations quantization, lossy compression.

D. *Feature Extraction*

Generally, feature selection can be divided into two groups: filtering method and wrapper methods. Filtering methods do not include learning and from learning classifiers they select features independently. The drawback of filtering methods is that they doesn't consider the possible interaction of features among them and consider single features in isolation. Although, the combination of these features may have a consolidate effect which doesn't follow from the separate interpretation of features in the group. Whenever, there is a limitation on the number of features to be selected, efficient features may not be included. The wrapper methods is based on feature subsets in terms of estimated classification errors and then build the final classifiers. Moreover, performing feature selection in steganalysis offers some advantages.

1. Eliminating the incomprehensible features.
2. Enhancing the performance of classification.
3. Minimizing the complexity of both feature generating and classifier training.
4. To find out the features that are fragile to steganographic method.

The generation of feature vector from the image database can be done by analysing from four efficient and famous steganalyzers Pevny-Fridrich, Chen, Lyu-Farid, YASS. Then these features are normalized into the unit interval[0,1] to use the identical membership function in the fuzzy rule generation. Mainly histogram of spatial representation and discrete wavelet representation features are considered, because of its statistical artifacts on the images. The features such as Discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. The key advantage of wavelet transform is that it captures both frequency and location information (location in time). Mainly three types of wavelet of the images are obtained the Haar wavelet, the Db4, and Symlet wavelets.

E. *Evolutionary Fuzzy Systems*

Fuzzification is the process of conversion of a precise or a extact value quantity to a fuzzy quantity that is variable. Each and every fuzzy if-then rule generated is coded as a string and denoted by six semantic values don't care(DC), small(S), medium small(MS), medium (M), medium large(ML), and large(L). The rectangular membership function represents the DC, and its value is 1. CF denotes the certainty factor of the rule. Each fuzzy rule has a certainty factor that describes the strength of the rule of its part. The membership function of the above semantic value in Fig 3 is described by using triangular membership functions. Though, we can use other membership functions, the total number of possible fuzzy rules would become 6^n (as there are six semantic values) such as with n-dimensional feature vector. It is not practical to impart all those fuzzy rules in a single fuzzy rule base for huge value of n. Therefore, our evolutionary method finds high performance fuzzy rules. Performance means the ability to demonstrate the pattern of stego images.

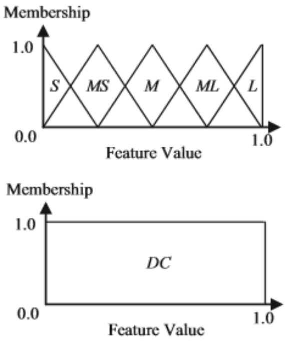

Figure 3. Membership Function

F. Evolutionary Fuzzy Algorithm

The involved evolutionary fuzzy algorithm optimizes one fuzzy rule in repeat execution of the algorithm and discover rules iteratively[4]. Firstly the same weight is assigned to all training samples and each instants of evolutionary algorithm is assigned corresponding feature vector of an image. In each steps of the algorithm, output of the iteration considers the if-then rule with the highest fitness value. Then the learning process minimizes the weight the samples which are learned correctly in the previous iteration. Samples with higher weight are considered more in the training process. Hence, the next cycle finds the fuzzy rules that are to be revealed for further. In short, the fuzzy rules that encorporate all the training samples than any other are added to final rule base. In this learning system,[5] a fitness function is used in the evolutionary process, which is computed for example for discovering pattern. fitness(R)= $_p$* f_p − w_n* f_n where f_p is the rate of positive training samples enclosed by rule R and f_n is the rate of negative training samples enclosed by rule R. w_p, w_n are the weights of positive rule and negative rule. The overview of the iterative evolutionary method is as described:

1. Initialization: Initialize an initial set of fuzzy if-then rules based on the weight.
2. Generation: Produce new rules by genetic operations.
3. Replacement: Substitute portion of current rules with the newly generated rules.
4. Inner Cycle Termination Test: Terminate the iteration on Step 2 and Step 3 if stopping criteria is met, else go to Step 2.
5. Outermost Cycle Termination Test: Terminate the iteration from Step 1 to Step 6 if stopping criteria is met, else go to Step 6.
6. Weight Adjustment: Minimize the weight of samples that encloses the new fuzzy rule obtained during algorithm. Go to Step 1.

The output of each outer cycle of the above algorithm is exactly one fuzzy if-then rule. Therefore, in the iterative learning, partial solutions are repeated no of times to obtain the complete solution.

IV. RESULTS AND DISCUSSIONS

The experiments were executed on a personal computer with Pentium processor, and Matlab R2013a was used for program writing. Images to implement this method was obtained from Washington University image database[11]. Then these images were converted to Grayscale images. These images were all encrypted to obtain the stego images by employing steganography methods. Data's were embedded into images by using MB, YASS, PQ methods. Following that, obtained the features of the encrypted images. Mainly the histogram of spatial representation and the discrete wavelet transform representation were considered, since it can be used to detect the statistical artifacts of the images. The use of Evolutionary fuzzy system as a rule generation tool is because of its ability to produce correct and intelligible knowledge. Whereas in other learning systems like Support vector machine(SVM), Naive Bayes, K-nearest neighbours the final knowledge base does not include the efficient features from the input. Moreover, in evolutionary fuzzy systems one can understand the produced knowledge base based on fuzzy rule. It is also possible that the learning process is done as an optimization problem, and can be solved by EA.

By the use of fuzzy system, those semantic rules which represent the decision processes makes more understandable by humans; but fuzzy systems are suffering from difficulty of self learning and documentation of knowledge. However; genetic algorithms provide strong search capabilities in complex global and local spaces whereas fuzzy systems present flexible inference methods to incorporate with incorrect and uncertainty. Hence, to achieve the advantages of both hybridization of FL with GA becomes essential. This can done by obtaining the signature of stego and clean rules as fuzzy rules using triangular membership functions. Further Evolutionary fuzzy algorithm is implemented by optimizing the fitness function. Simulations were carried out and uniform crossover is used for performing crossover and selection was done based on tournament selection mechanism to find the best individual.

Table1: Values of parameters in simulations

Parameter	**Value**
Population Size	100
Don't care substitution rate	0.5
Crossover probability	0.9
Mutation probability	0.5
Fitness positive weight	0.1
Fitness negative weight	0.9
Substitution percentage	10
Maximum no of iterations	50

On estimating the evolutionary method, the complexity in extracting fuzzy if-then rules by an evolutionary iterative rule learning is high. The complexity of genetic algorithm using fuzzy would be in a quadratic form of 0(n^2). But inorder to detect secret data in an image, only matched fuzzy rules are checked so that the complexity in extracting rules can be tolerated. Since simulations were done on gatool tool, steganalysis fitness function was optimized faster for specified no of generations. The inference obtained on using Evolutionary fuzzy rule as a learning compared to other learning is mentioned as follows

1. Fuzzy logic have the ability to deal with uncertainties such as imprecision and doubtness involved usually in classification problems ie, here the stego is represented using six semantic values as small(s), medium small(ms), medium(m), medium large(ml), large(l), don't care(dc) using membership functions to represent it.
2. Samples can be categorized by a set of fuzzy rules based on the semantic values of its attributes associated with it.
3. The major ease of GA is that it can optimize values from very huge search space and able to learn.
4. Fuzzy decision making approach can been employed to handle inaccurate information.
5. Evolutionary fuzzy system to use as a rule generation tool is due to its ability to generate correct and intelligible knowledge.
6. When manual selection of their values becomes difficult GA can be used to optimize the value.
7. Genetic Algorithms produce strong search capabilities in huge spaces and fuzzy systems present ease inference methods to incorporate with inaccurate and unreliability.
8. Thus the incorporation of GA and EFS can be done by taking the advantage of GA's learning and the advantage of fuzzy systems which provides the adaptable inference methods in order to handle with inaccurate and unreliability together.

Thus through the hybridization of genetic algorithm and evolutionary fuzzy system, optimized the fitness function for about 50 iterations.

V. CONCLUSION

Thus through evolutionary fuzzy systems and genetic algorithms, stego images can be detected from clean images. Basically, the approach consists of two stages: Firstly, an evaluation of training set to explore the unique markings of stego images, and next, the steganalyzer is prepared to find out the steganography method.

In other words, to find the stego images, after exploring the used steganography method, an appropriate model for steganalysis was employed and can be improved further by detecting hidden message, embedded key estimation and message length etc. In this aspect, an evolutionary fuzzy algorithm is proposed to generate fuzzy rules from features of the stego images. These rules are used to form the pattern of steganography methods. In line with the obtained results, this approach involves evolutionary method in contrast to the plain use of steganalysis methods. The major benefit of this viewpoint is that on the arrival of modern steganographic method, the fuzzy rule base can be enhanced. In future it can be also extended to extract the hidden information of the images using cryptographic techniques. Thus it can be implemented in the areas where security is the main concern.

REFERENCES

[1].HediehSajedi, "*Steganalysis based on steganography pattern discovery"*, journal of information security and applications, Vol. 30, pp. 3-14, 2016.

[2].David Garcia, Antonio Gonzalez, Raul Perz, "*A feature Construction approach for Genetic Iterative Rule learning Algorithm*", Journal of Computer and System Science, Vol. 80, pp. 101-117, 2014.

[3].Archana O. Vyas, Dr. Sanjay V. Dudul, *"Study of Image Steganalysis Techniques",* International Journal of Advanced Research in Computer Science, Vol.6, pp.7-11, 2015.

[4].Xiangwei Kong, Chaouy Feng, Ming Li, Yanqing Guo, *"Iterative multi-order feature Alignment for JPEG mismatched Steganalysis"*, Neuro computing, Vol. 6, pp. 1 - 13, 2016.

[5].Daniel-Lerch Hostalot, David Magias, *"Unsupervised Steganalysis based on Artificial Training sets"*, Engineering Applications of Artificial Intelligence, Vol. 50, pp. 45-59, 2016.

[6].Kaushal Solankitt, Anindya Sarkart, B.S.Manjunath, *"YASS: Yet Another Steganographic Scheme that Resists Blind Steganalysis"*, 2017.

[7].Phill Salee, "Model Based Steganography*"*, Proc. Conference on Computer Science, 2016.

[8].Jessica Fridrich, Miroslav Goljan, David Soukal, *"Perturbed Quantization Steganography with Wet Paper Codes"*, Conference on ACM, 2004.

[9].Huai-xiang Zhang, Bo Zhang, Feng Wang, *"Automatic Fuzzy Rules Generation Using Fuzzy Genetic Algorithm",* Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009.

[10].Shyi-Ming Chen, Fu-Ming Tsai*, "Generating Fuzzy Rules from training instances for fuzzy classification systems",* Expert Systems with Applications, Vol. 35, Issue. 7, pp. 971–987, 2002.

[11].http://www.cs.washington.edu/research/imagedatabase

[12].Der-Chyuan Lou, Chao-Lung Chou, Hao-Kuan Tso, Chung-Cheng Chiu, "*Active steganalysis for histogramshifting based reversible data hiding*", Optics Communications, Vol. 285, pp. 2510-2518, 2012.

[13].R.Chandramouli, *"A mathematical framework for active steganalysis",* Multimedia systems, Vol. 9, pp. 303-311, 2003.