Cognitive Prediction of the Most Appropriate Image Steganography Approach

Usha B.A
Assistant Professor
Dept of CSE, R.V.C.E,
Bangalore - 560059

N.K Srinath, PhD
Prof and Dean, PG Studies, Dept of CSE, R.V.C.E,
Bangalore - 560059

Ravikumar C N, PhD
Prof and Head,
Dept of CSE, SJCE,
Mysore - 570006

Vismaya S P
Final Year BE,
Dept of CSE, R.V.C.E,
Bangalore - 560059

ABSTRACT
In the light of growing technological advancements, data security has become a matter of great concern. This in turn has reiterated the significance of fields like cryptography and steganography to modify or hide secret data. The conventional methodologies of steganography and cryptography are not designed to utilize the semantic meaning of data. Even though these traditional methods offers ample security, it fails to personalize the whole process and make it dependent on the requirements as specified by an individual. Thus the idea of cognitive cryptography took birth. The field of cognitive science has been progressing at a fast rate and deals with decision making. In this paper, the idea of cognitive cryptography has been adapted to give rise to cognitive steganography. The application developed here aims at deciding the most suitable steganography approach (among the chosen four algorithms) for hiding input data, by taking into account its semantic meaning and the intended application. The development of such an algorithm overcomes the shortcomings of traditional techniques in terms of computational complexity, memory usage, image distortion and effective bandwidth utilization.

Keywords
Cognitive Steganography, Artificial Neural Networks, Text Mining, Machine learning techniques Image Steganography

1. INTRODUCTION
In this world where technology is growing ever so widely at an exponential rate, the need for data security is highly indispensable. There are software and techniques developed to secure data and in the same way techniques which are twice that number are being developed to attack the data and obtain the information. For this very reason we should use data hiding techniques which deliver the following:

- Potential ability to conceal the presence of private information
- Difficulty of detecting the hidden (i.e., embedded) data
- Ameliorate the encrypted data's secrecy

Steganography is the ever growing science in field of cyber security. It is the art of hiding data of any type like text, image, audio or video in a cover medium like text, video, audio or image. [1] The steganographic process is as shown in Fig 1.1:

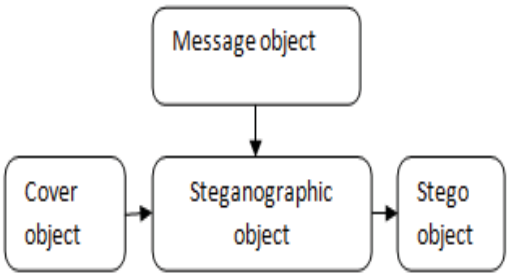

Fig 1.1 Steganographic Process

Data hiding techniques have constantly been an area for probable improvements like employing hybrid techniques [2] combining different approaches. Also enhancement of steganography techniques is by employing some more levels of encryption and hiding to base algorithm [3]. Cognitive systems [4] are an upcoming branch of computer science which attempts to bring a human touch to decision making process. Traditional steganography fails to provide a personalized approach to the whole data hiding process since it gives a generic solution to the problem. Now that, steganography is collaborated with cognition, much importance is given to the semantic meaning [5] of data in question. Depending on the security requirements particular to the application, the decision on the kind of steganography is made. This helps in reducing the overhead encountered in the computation and effectively the time complexity.

This paper aims to determine the most suitable steganography algorithm to hide the given input data in a cover image, using cognitive science. Cognitive sciences are an upcoming branch of computer science growing for many years and are related with fields such as psychology, neurobiology, linguistics, and recently technology [5] .The methods of cognitive science, or in other words, artificial intelligence, are most preferred in applications which are too complex to be programmed manually by a human being and in situations which require the system to customize and adapt to its environment after deployment.

1.1. Fundamentals of Machine learning
Machine learning is a branch of artificial intelligence which deals with the construction and study of algorithms which learn from data. There are a number of machine learning algorithms of which neural networks is a popular one, and which will be implemented in this project as well. Machine Learning is a scientific discipline that can program systems to automatically learn and to improve with experience.

Based on the data the machine learns by recognizing patterns which are complex and finally make intelligent decisions. Though there is difficulty in describing all the possible sets of decisions made from all the possible inputs given since they are complex. This problem can be tackled by Machine Learning which is a search for algorithms obtaining knowledge and experience from specific data and experience that build classifiers in terms of predictor features. As a result of the training process, we get a reusable model [6] to which, during the use phase, new inputs that the system has never seen during the training phase can be submitted, then providing an output based on past experience.

ML converts data sets into pieces of software, known as "models," that can represent the data set and generalize to make predictions on new data. ML can be used in three different ways:

- **Data Mining:** ML can be used to gain insights from large databases.
- **Statistical Engineering:** ML can be used to convert data into software that makes decisions about uncertain data.
- **Artificial Intelligence:** ML can be used to imitate the human mind, to create computers that can work like the human brain in perceiving and seeing the working of various things around us.

People can understand complex structures if they relate to more isolated yet understandable concepts. Despite this fact, popular pattern recognition tools, such as decision trees or production rule learners, produce only flat models which do not build intermediate data representations. On the other hand, neural networks typically learn hierarchical but opaque models [7].

Steganography means covered or secret writing which comes from the Greek words *steganos* meaning "covered" and *graphein* meaning "writing" [8]. The goal of steganograpgy is to hide messages in such a way that no one apart from the intended receiver knows that a message has been sent. This can be achieved by concealing the information in harmless carriers or cover signals.

Supervised learning (machine learning as shown in Fig 1.2) takes a known set of input data and known responses to the data, and seeks to build a predictor model that generates reasonable predictions for the response to new data.

Fig 1.2 Supervised Learning

2. IMPLEMENTATION OF COGNITIVE SYSTEM AND STEGANOGRAPHY

The implementation phase in any project development is the most important phase as it yields the final solution, which solves the problem at hand. The implementation phase involves the actual materialization of the ideas, which are expressed in the analysis document and developed in the design phase. Implementation should be a perfect mapping of the design document to a suitable programming language in order to build the desired system. Often the product is ruined due to the selection of incorrect programming language for implementation or unsuitable method of programming. It is better to directly link the coding phase to the design phase in the sense that, if the design is in imperative terms, then the implementation should also be preferably carried out in an imperative manner. The factors concerning the programming language and platform chosen are described in the forthcoming sections.

The implementation stage is a system project in its own right. It involves

- Careful Planning
- Investigation of the current system and the constraints on implementation.
- Training of staff on the newly developed system.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Evaluation Metric

The main metrics in this project are the MSE and PSNR values calculated based on the user security requirements, payload capacity, invisibility and robustness of the given file depending on which the most suitable image steganography algorithm is selected.

3.2 Experimental dataset

The data set used in our project is both obtained online and created which have 100% genuine content. Most of them are text files with images of the hospital logos or diagrams in question papers which are strictly in PDF format.

The datasets include medical reports financial records, question papers and weather reports under which there are 14 sub categories.

3.3 Mapping of the performance parameters to the steganography algorithms

Any steganography algorithm is evaluated on the basis various parameters like invisibility, Payload capacity, Robustness against statistical attacks, Robustness against image manipulation Unsuspicious files.

Since this application aims to choose the most suitable steganography algorithm for the required purpose, it was necessary to map these performance features to the four steganography algorithms chosen. Which steganography algorithm satisfies which of the above criteria and with what efficiency was an important consideration, keeping in mind the application it is intended to. For Example, in case if the predicted algorithm by the ANN is Spread Spectrum then its stego and cover images are displayed along with PSNR and MSE values and a comparative chart (shown below) is displayed for user to see PSNR and MSE of other algorithms in consideration as shown below in table 3.1:

Table 3.1 Performance comparison of chosen steganography algorithms

Stegnography Algorithm	Payload	Invisibility	Robustness
Least Significant Bit	Low	Medium	Low
Least Significant Bit with XOR Encryption	Low	Medium	Medium
Spread Spectrum	Medium	High	Medium
Gray Code	High	High	High

3.4 Performance and Result Analysis

To check the accuracy of the application and to analyze the results obtained for a given input we are showing two testing cases, one using default options and one with the user requirements is the first priority.

The first set of results were obtained using default options suggested by the application. Here the user uploaded a simple requisition form provided in the hospitals which general information of the patients. As a requisition form requires less security its payload is less, invisibility is less and robustness is also low and for this combination the system is trained to select an algorithm which offers lesser security.

The Performance of all the four candidate algorithms are illustrated below with respect to size of different text files as shown in Fig 8.1, Fig 8.2, Fig 8.3, Fig 8.4 and Fig 8.5. In the graphs X-coordinate specifies different file sizes and y-coordinate shows corresponding value of MSE. In Fig 8.1, the accuracy of the document classifier in categorizing input documents into one of the chosen fourteen categories is shown. The accuracy depends on the frequency of the words in the document along with the context in which they occur. It also depends upon the presence of images, which are not extracted, resulting in incorrect of failed document classification.

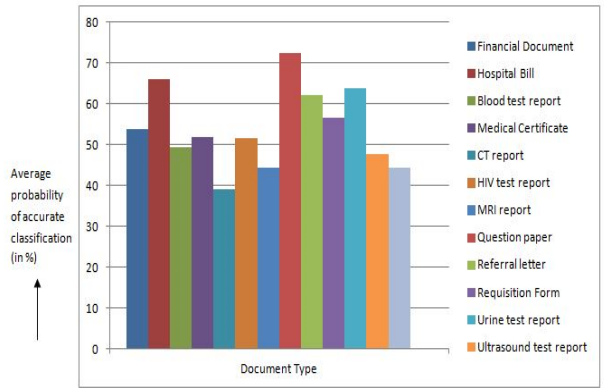

Fig 8.1 Accuracy of the Document Classifier

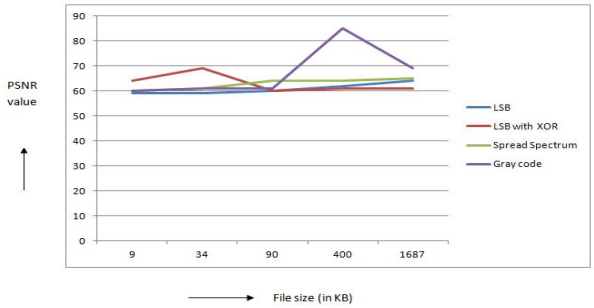

Fig 8.2 Behavior of candidate algorithms with different input file sizes

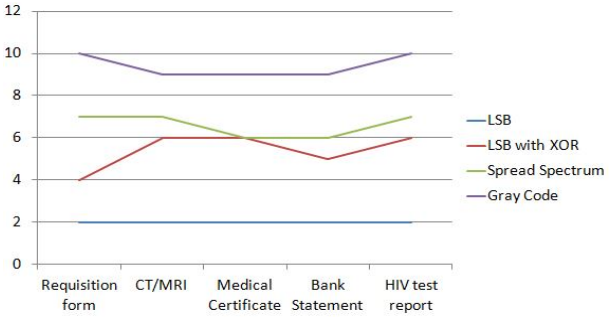

Fig 8.3 Behavior of candidate algorithms with respect to document types having different security requirements

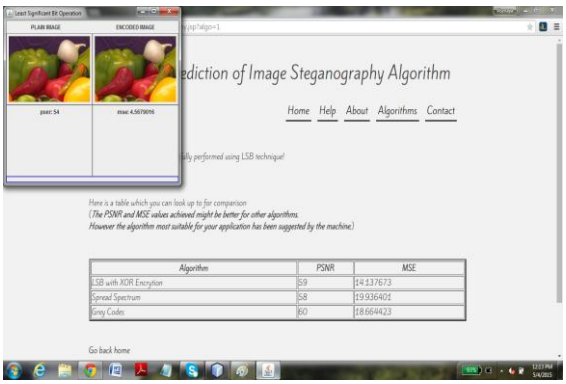

Fig. 8.4: Results with default values of payload, invisibility and robustness

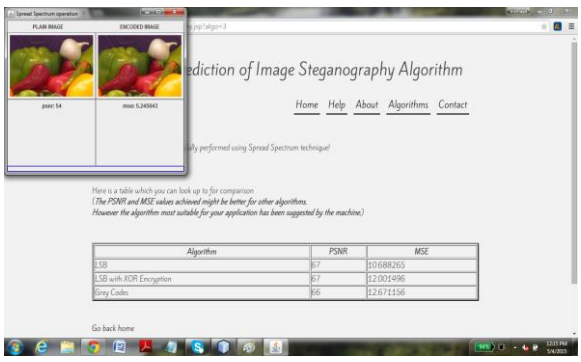

Fig. 8.5: Results when security level is mentioned by the user

4. CONCLUSION AND FUTURE ENHANSEMENT

In a world where data thefts and attacks are increasing the need for data security is highly required. This project aims at suggesting the best suitable algorithm for a given data based on vectors like payload, invisibility and robustness. The project works accurately after the required steps which have been mentioned have been carried out successfully.

The document classification and steganography methods produced expected results. The steganography algorithms which are used are LSB algorithm, LSB with XOR algorithm, Spread Spectrum algorithm and Gray Code algorithm. The predictions made by the artificial neural network showed an accuracy of about 87.2% in the considered dataset. The project can be extended to include all kinds of multimedia inputs like images, audio and video. Accuracy plays a very important role while predicting the steganography algorithms. Models with higher accuracies can be developed where the neural networks can show accuracy close to 100% by increasing the number of datasets and training the machine extensively with more number of features such as specific characteristics of the stego-image. This project can be extended to retrieve back the information hidden in the stego-image in its original format (in PDF). A document classifier with better efficiency can be implemented.

5. ACKNOWLEDGMENT

It is our privilege to acknowledge thanking all the department personals and sponsors who gave us an opportunity to present a paper at this level. We wish to place our deep sense of gratitude to all reference papers authors for their beneficial papers, books and websites etc

6. REFERENCES

[1] Srinath N K, Usha B A, Narayan K, Tushara C K, “Analysis of Data Embedding Technique in Image Steganography – A Survey” , in International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 6, pp.54-59, June 2014.

[2] Priyadarshni, “A hybrid data hiding scheme to enhance the capacity of one-third probability embedding method”, in 2015 IEEE International Conference on Computational Intelligence & Communication Technology, ISBN: 978-1-4799-6022-4, Pages: 269 – 272, 13-14 Feb. 2015.

[3] Ramandeep Kaur, Abhishek Thakur, Hardeep Singh Saini, Rajesh Kumar, “Enhanced Steganographic Method Preserving Base Quality of Information Using LSB, Parity and Spread Spectrum Technique”, 2015 Fifth International Conference on Advanced Computing & Communication Technologies IEEE,Pages:148 – 152, ISBN:978-1-4799-8487-9, 21-22 Feb. 2015.

[4] Marek R. Ogiela, “New Directions in Cognitive Cryptography”, in Journal of Convergence, Volume 5, Number 3, September 2014.

[5] Lidia Ogiela and Marek R. Ogiela, “Towards Cognitive Cryptography”, in Journal of Internet Services and Information Security (JISIS), volume: 4, number: 1, pp. 58-63, 2014.

[6] Chin-Teng Lin, Fellow, Mukesh Prasad, and Amit Saxena, “An Improved Polynomial Neural Network Classifier Using Real-Coded Genetic Algorithm”, Systems, Man, and Cybernetics: Systems, IEEE Transactions on (Volume: PP, Issue: 99), ISSN :2168-2216, 12 March 2015.

[7] Jan Chorowski and Jacek M. Zurad, “Learning Understandable Neural Networks with Nonnegative Weight Constraints”, IEEE Transactions on Neural Networks and Learning Systems, Vol. 26, No. 1, January 2015.

[8] Masoud Nosrati, Ali Hanani, Ronak Karimi, "Steganography in Image Segments using Genetic Algorithm ", in 2015 Fifth International Conference on Advanced Computing & Communication Technologies, ISBN:978-1-4799-8487-9,21-22 Feb. 2015

[9] Kirti Bala Bahekar, Praneet Saurabh and Bhupendra Verma, ―Improving High Embedding Capacity Using Artificial Immune System: A Novel Approach for DataHiding, in Proceedings of All India Seminar on Biomedical Engineering 2012 (AISOBE 2012)Lecture Notes in Bioengineering 2013, pp 209-219, 02 Nov 2012

[10] Mamta juneja, Parvinder Singh Sandhu , ―Improved information security using Steganography and Image Segmentation during transmission‖,seminar in Rayat and Bahra Institute of Engineering and Technology (RBIEBT) ,2011.

[11] Khalid A. Darabkh, Iyad F. Jafar, Raed T. Al-Zubi, and Mohammed Hawa, ―An improved Image Least Significant Bit Replacement Method. MIPRO 2014, 26-30 May 2014, Opatija, Croatia.

[12] Mrs. Sayantani Ghosh, Mr. Sudipta Roy, and Prof. Samir K. Bandyopadhyay, ―A tutorial review on Text Mining Algorithms‖, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 4, June 2012.

[13] Niti Syal, Naresh Kumar Garg, ―Text Extraction in Images Using DWT, Gradient Method and SVM Classifier, International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 6, June 2014.

[14] Vishal Gupta, Gurpreet S. Lehal, ―A Survey of Text Mining Techniques and Applications‖, Journal of Emerging Technologies In Web Intelligence, Vol. 1, No. 1, August 2009.

[15] Divya Nasa, ―Text Mining Techniques- A Survey, International Journal of Advanced Research in Computer Science and Software Engineering Research Paper, Volume 2, Issue 4, April 2012