12

Fuzzy Logic on a Polygenic Multi-Agent System for Steganalysis of Digital Images

Samuel Azevedo, Rummenigge Rudson
and Luiz Gonçalves
Universidade Federal do Rio Grande do Norte,
DCA-CT-UFRN, Campus Universit[ario,
Lagoa Nova, Natal, RN,
Brazil

1. Introduction

Digital cryptography has being a solution for protecting transmission of data in applications such as electronic commerce (Luciano 2003), electronic vote (Kofler 2003), and digital Television (Macq 1995). However, an interceptor monitoring network flow could easily break purely encoded data and clear the contents of cryptographed messages. Steganography techniques came up in order to help improving this protection. The goal of steganography is to hide data into a covering message (envelop) in such a way that an interceptor has no way to notice the presence of a hidden message in its covering envelop. Note that one can combine both cryptography and steganography in order to achieve better security. For example an image can be enriched with visually imperceptible extra information that, when eventually noticed, could be understood as an eventual noise. This damaged image could serve thus as a camouflaging body that brings protected data to the other side of the communication process. Any media object can be used as the covering message, such as text, audio, video, network packages, and file systems. Digital images are known to be the most used media objects for this purpose due to its inherent artistic appeal.

In steganography, specifically, the carrying message is a digital object (image, audio, video etc) that envelops hidden data. When a potential covering object carries hidden data it can be called a steganographed object. In order to extract hidden information from this object, one has to know that a conspicuous object is steganographed, what is the steganographic algorithm used to hide data, and the password that will be generally requested by the algorithm.

On the other side, if one would like to reveal the data which is hidden, it should use steganalysis techniques in order to detect whether a message has hidden data or not. This is just the subject approached in this work. Although it is undeniable that everyone has the right to protect some information there are some situations where it is necessary to reveal its contents, for example breaking of privacy is important in criminal investigations.

Besides detecting the presence of hidden messages, a useful steganalysis technique should also estimate the length of the messages and also somehow possibly to detect which steganographic algorithm is used to hide information. Since one knows the algorithm to

hide or to reveal the message contents, and the steganographed object, one may try some common known attacks to break the password like the brute force password guessing.

When new cryptographic or steganographic techniques arise, new cryptoanalysis or steganalysis also are developed addressing the new characteristics of the problem. So, one can say that there is a race between cryptographers and cryptoanalyzers and between steganographers and steganalysers. A technological advance in one side forces the other to overcome it.

Other characteristic of the problem is that when new steganalysis techniques are developed, new steganographic techniques arise immune to the existent attacks. Therefore steganalysis systems demand flexibility to adapt to the new steganographies. This flexibility can be obtained by learning or by using software engineering techniques that ease the alteration of the system in a handful time (such as modularization, documentation, etc).

In this work we approach steganalysis for digital images, which represent a vast distribution of data around the Internet. Due to the very complex nature of the problem, it is generally required to perform steganalysis on a huge volume of data. Of course it would be adequate to perform this in an autonomously way by using a computational system. Autonomy and flexibility are characteristics present in software entities called agents. By the complexity of the problem, these agents would be more appropriately approached in a Multi-Agent System (MAS), which is a system where several specialized or redundant agents interact (through cooperation, negotiation, and exchanging information, for example) to achieve their goals.

Since MAS are systems that approach social interaction between agents, we need to model the way these interactions will be performed. It is common to use metaphors from nature as heuristics in order to solve computational problems in a less complex way. A good heuristic for this solution would be inspired in social interaction of insect communities. Social insects present important characteristics of MAS such as cooperation, distribution of multiple tasks, and coordination. Our work is inspired in the polygenic societies of bees from the species Melipona Bicolor where several queens of a hive can cooperate in the coordination of all the workers. We initially apply such coordination model to our MAS, where each worker is a classifier, and further apply fuzzy logic to solve the classification of heterogeneous classifiers to a same sample.

Therefore, our proposal and main contribution is a multi-agent system for digital image steganalysis that is based on the paradigm of the community of polygenic bees using fuzzy logic. With such approach we aim to solve the problem of automatic steganalysis for digital media with a case study on digital images. The architecture proposed here is designed to detect if a file is suspicious of carrying hidden contents allowing to attempt to extract them with other techniques (such as brute force password guessing). Experimental results validate the system, showing the applicability of the MAS to steganalysis of image data.

2. Background and methods

In order to better understand our problem, some background must be addressed in different areas of knowledge including cryptology, machine learning, MAS, heuristics, image segmentation, and fuzzy logic.

2.1 Machine learning

"Machine Learning is the AI field which aims to develop computational techniques about learning as well as the construction of systems capable to acquire knowledge in an automatic way." (Rezende 2003)

Among the machine learning paradigms, we have (Sanches 2004):

- *Symbolic paradigm* – builds a symbolic representation of the problems´ solution through the analysis of examples, the machine learning most known methods of this paradigm are *decision trees* and *semantic networks*.
- *Statistical paradigm* – composed by the classification methods that try to analyze statistics in order to find an statistical model approximated to the problem; a known method of this paradigm is the *Bayesian Learning* algorithm.
- *Paradigm based in examples* – classifies one instance (or sample) through its comparison with other previously classified samples, returning as result the class of the classified instance that is more similar to it; the most known method of this paradigm is the K-nn which returns the class that appears the most in the *k* nearest neighbors to a consulted sample.
- *Connectionist paradigm* – based upon the biological metaphor of neural connections of the nervous system, it try to train a network of neurons with samples in a way that the weights of its connections are adjusted to solve the problem of classification.
- *Evolutionary or genetic paradigm* – this is also based in biological metaphor, in this case the genetic evolution; it consists in realizing crossings and mutations in a set of classifiers to solve a problem; during N interactions (or generations), the classifiers with best performance in each generation prevail and the next generation of classifiers is generated by variations of these; the genes are the parameters of the classifiers, that can be of any of the other paradigms, but instead of regular training to accurate they parameters, these parameters are changed through evolution.

Every learning method presents, after training, an error or accuracy rate. Other important rates are the True Positive and True Negative rates that indicate respectively the rates of positive and negative cases correctly detected. Many times, one wish to improve these rates and one of the improving strategies are the ensembles or clustering of classifiers. In steganalysis, a critical rate is the False Negative, which indicates the percent of cases that were incorrectly classified regular images but in fact contained hidden data.

2.2 Multi-agent systems

Agents are autonomous software entities that act in a certain environment and are capable of taking decisions as which actions to perform in order to reach any goal (Russel 1995).

A multi-agent system is a complex system in which several specialized or redundant agents interact, cooperating, negotiating, and exchanging information in order to reach any optimal goal.

MAS are systems that contain a set of software agents working together that interact between them and with the environment through some communication channel. Agents have areas of influence in the environment that may or not overlap (Wooldridge 2001). They

can interact through the use of negotiation, coordination, or cooperation. Bid, argumentation or game theory can also be used by agents (Macedo 2001).

A society of agents may be composed by homogeneous or heterogeneous agents. The coordination problem is how to manage the interdependencies of tasks and resources between agents. Wooldridge classifies four models of coordination: global-partial planning, joint intentions, mutual modeling, and social rules.

In our problem, we use a heuristic of social bees to coordinate the collective work of agents, and the we approach in this MAS a fuzzy clustering algorithm to enhance the detection of hidden data into images.

Fig. 1. M. Bicolor queens in reproduction process. Two or more bees can put much more eggs thus diminishing a lot the efforts for getting a mature colony.

2.3 Heuristics

Heuristics can be devised base on approaches as genetic algorithm, memetic algorithms, simulated annealing and insect colonies as ant and bees. Algorithms that use metaphors based on colonies aim to imitate some behavior of those in order to search solutions for complex problems. Biologically, social insects may be monogenic or polygenic. That means, it can exist societies that present a single or several queens at the same time (Aponte 2003). Bees of the specie Meliponine Bicolor (see Figure 1) can be polygenic.

2.4 Image segmentation

Since steganography aims to hide the existence of data within data, it´s important to find computer vision techniques that are able to see this hidden information in images. The most simple steganographic algorithms aim to hide data in the less significant bit of each pixel, these generally are imperceptive to human eye, but they generate distortions in images easily detected by common image segmentation algorithms.

Although there are general purpose techniques and algorithms for image segmentation, they often must be combined with domain knowledge to effectively solve a vision problem; thus, image segmentation must be approached by many perspectives (Pavidlis 1982).

Methods based in edge detection, histogram statistics and clustering, and transform domain error prediction, are found in many of the current solutions for steganalysis, as the discussed in section 2.6.

In our work, we use some of the methods above to compose the features that compose an instance for the machine learning algorithms. These features use statistical information such as mean, variance, asymmetry and kurtosis. Mean is the first statistical momentum, variance, asymmetry and kurtosis are, respectively the second, third and fourth momentums over the mean. The equation bellow shows the general formula to find the kth momentum (the mean is the first momentum, but its value is 0). The equation also can be read as *E[(X − E[X])k]* where *X* is a random variable, and *E[X]* is the expected value.

$$\mu_k = \left\langle \left(x - \langle x \rangle\right)^k \right\rangle = \int_{-\infty}^{+\infty} (x-\mu)^k f(x) dx.$$

2.5 Fuzzy logic

Since the publication of "Fuzzy sets" (Zadeh 1965), many studies have been done to apply fuzzy logic in diverse fields. In machine learning, fuzzy logic has been applied to algorithms from different paradigms as well as to ensembles of classifiers, for example: Support Vector Machines (Lin 2002); neural networks (Carpenter 1992), (Jang 1993); and decision trees (Acampora 2011). López-Ortega (2011) points out that fuzzy clustering and MAS lead to high quality decisions.

In software agency, we can see the use of fuzzy knowledge based systems (Arroyo 2011) to implement the decision making process and actions of software agents. Also, we can observe the use of fuzzy logic theory for agents coordination (Goodarzi 2011), (Hagras 2010).

In steganalysis, the most common use of fuzzy logic is presented in the use of fuzzy machine learning algorithms and in fuzzy clustering (see the related work in 2.6 for further details). One of the main contributions of this work is the design of a novel fuzzy clustering approach using coordination of agents.

2.6 Cryptology

Steganography is a subarea of information security that includes several other inner areas meaning covert written (Katzenbeisser 2000). In general, its focus is the inclusion of information in a media data that is not suspect. In fact, it is the art of occluding data in data (Artz 2001). When two communication sides A and B want to exchange a secret message, they use an occulting message (or covering object, envelop, mule) applying some steganography technique that may use or not some key *k* obtaining in this way a steganographic message that is undistinguishable from the previous. This last is sent through the communication channel. There are several techniques for doing steganalysis as:

1. S**ubstitution system –** redundant parts of the media are substituted by the data that one wants to occult;
2. T**echniques in the transform domain –** insert secret data in the signal transform domain (frequency domain);
3. S**pecter scattering –** the specter of distribution of the information is scattered;

4. **Statistical methods** – produces steganographed data through statistical manipulation of covering data;
5. **Distortion techniques** – produces distortions in a covering media in order to get steganographed data, compares the original covering media with the modified in order to extract them.

An important characteristic in steganography is determining the capacity of an object to hide information, we can observe this concept in what Moskowitz (2002) calls Capability:

"Capability = (P;D) where P is the payload size and D is a detectability threshold. We sometimes expand the capability to a triple (P; D; R) where R is a measure of robustness of the stego channel."

The quoted author also states there for steganography in the least significant bit of images, the payload is limited from 0 to 50% of the size of the carry image, otherwise changing the cover to a negative.

Steganalysis goal is to attack or monitor a communication channel in order to detect existing information that is occulted in messages or to forge some occult message, interrupt communication, and to extract occulted data.

Different approaches to steganalysis can be found in the literature as visual attack (Fridrich, 2002, 2004), statistical analysis (Katzenbeisser 2000), and signature detection (Chandramouli 2004). The first approaches the most elementary methods, as for example the bit substitution systems that may cause visible distortions to images, what reveals the existence of hidden contents. Statistical analysis looks statistical measures in files as the histogram to verify common aberrations. It can use pure statistic methods or some combination with machine learning. In signature detection approaches, any degradation caused by steganography methods can be read as a signature of these methods. These methods generally span suspect files to find signatures in the data noise that can reveal if any steganography approach is used including some times which was the used approach.

There are two categories of steganalysis techniques: specific and universal (or blind) steganalysis. While specific steganalysis is related to attack objects generated by one single steganographic algorithm, universal steganalysis aim to attack stego objects independent of the steganographic algorithm used. Commonly, steganalysis use machine learning algorithms in order to classify whether an object may contain hidden data or not.

In order to create a steganalysis algorithm, one must think in six phases or steps:

1. **Steganalysis goals** - Consists in defining and implementing the category of technique will be performed (specific or universal), and defining which attacks will be realized, as detection of hidden information, data estimation (as for example the length of the hidden data), steganographic algorithm used. The following types of attack don´t need a classifier, and so, if they are isolated attacks it´s not needed to implement the steps 2 to 6, but if combined with the other attacks mentioned above, these steps are still necessary: data extraction (as password guessing from a dictionary), intercept the cover messages (such as sniffing network packages), denial of service (applying noise to an image, disabling the possibility to extract hidden content), and forging a hidden message to confound the communication.
2. **Classifier Method**- one must choose and implement which machine learning algorithms will be used to the classification process. If more than one classifier will be

used, also a clustering technique must be defined to combine the classification results of different algorithms. Sometimes, the architecture of the final algorithm must be redesigned and applied to new trainings and tests in order find improved results. According to the complexity of the features and data that will be analyzed, one may choose a most fitting machine learning solution. This can be performed before features calibration and obtaining data samples, or after these steps.

3. **Features Calibration** - one must select the features used to describe the data that will be applied to the machine learning technique; in the case of images, these features may be statistical data from histogram or from segments of the image, errors found in predicted coefficient values in the transform domain, and so on. This can be achieved by choosing an initial set of features (by literature, empirical experience, experimentation, etc) and testing subsets of these features in the next phases to verify the optimal subset of features. Liu (2008b) describes an interesting methodology for feature mining for steganalysis.
4. **Data Samples** - it´s necessary to create a database with samples fitting the features selected, and this database should able to be accessed by the classifier. But two random subsets of samples might be separated, the bigger to the training and smaller to the testing phases. The size of these databases is another issue, the ideal, statistically speaking is that this size should be big enough to represent the population of real cases; but by the nature of the problem, there are no statistics describing how many stego objects are there in the world; so, there are works using from 30 to more than 30000 samples. It´s important to say that the samples must be in quantities proportional to the different classes (from non-stego objects/stego objects in the most simple cases; to non-stego objects/stego object for algorithm 1…N in most complex solutions).
5. **Training** - this phase is about training the machine learning mechanisms with the training subset of data, according to the algorithm selected, this step may last a long time.
6. **Testing** - finally, after trained, the classifier may be submitted to the testing dataset, and the accuracy, true positive and true negatives rates must be calculated.

Bakhshandeh (2009) presents a steganalysis technique based on local information and human visual system. By performing segmentation and analysis for clustering these segments, the best segments are chosen for steganalysis. The algorithm they have used for classification is Fuzzy Clustering, simplifying, one may say that they give a fuzzy weight to the results of many classifiers, and use a clustering algorithm to decide the final classification results. Wavelet information is extracted to compose the feature used in a SVM algorithm to classify whether an image has or does not have hidden data. The results are at first sight promising, but if one consider that their experiments were in images carrying hidden data in 100% of theirs spectrum capability for spread spectrum steganography, one would expect to see the results for messages that are smaller the full capability of the cover image, since it´s more difficult to detect the presence of smaller data because the resulting alteration is smaller in the cover image data.

In the work of Liu and Sung (Liu 2008) it is presented a steganalysis technique that uses One-Against-All decomposition for SVM to classify whether or not a jpeg image contains hidden data in one of three steganographic techniques, based in detecting errors from predicted DFT, DCT or DWT coefficients. After this classification, they use a Dynamical Evolutionary Neuro-Fuzzy Inference Systems (DENFIS), to estimate the length of the hidden

data. The estimative found was very accurate for F5 steganographic algorithm, but not so effective for others.

Amirkhani (2011) highlights that blind steganalysis algorithms use to have each internal similar (or a same) processes for different image categories (smooth, complex, noisy, etc), instead of using the particular characteristics of an image type to attack it. Their framework can make use of any steganalysis technique that are applied to two main modifications: before training, the images must be divided into different content classes; and the result of a classifier must be weighted to a fuzzy value according to the content class trained, after that, the result is combined in order to classify if an image is a regular image or has a hidden content, these two final classes are called by the authors of Cover (regular) and Stego (has hidden content). They experiment this framework with some known steganalysis algorithms and confront their efficacy with several steganographic algorithms, showing discrete increases of accuracy, true positives and true negatives rates. In our approach, we train some of our classifiers for different image types, and other for general image types, in order to further clustering their results.

3. Polygenic MAS fuzzy clustering steganalysis

The MAS system approached here, according to the taxonomy presented by Rezende (2003), may be classified as a heterogeneous agent open system, with low initial granularity.

The main issue is social resolution that aims to solve the problem of steganalysis in a cooperative and distributed way. However, it is also approached the social simulation view for simulating the behavior of polygenic bees. Interaction patterns present in the system are commensalism (in the interactions between classifier and coordinator agents), and proto-cooperation, in interactions between classifier agents.

3.1 Architectural view

Figure 2 presents the general architecture of the proposed solution. The steganalysis process is realized in 2 steps: first, it´s necessary to perform some type of data interception (such as network packets sniffing) – this is not approached in this work; then, the intercepted data may finally be classified with our approach into stego data (data that carries hidden content) or non-stego data (without hidden content).

The polygenic heuristics here is present in the coordinator agents, which can represent the queens of this society. They are responsible to ask for specialized and general classifiers agents to analyze an intercepted image file. These queens also perform the following fuzzy clustering approach:

- According to the training, an **specialized classifier agent** may be more suitable to an image than other, so, it receives a bigger weight when classifying an image which category it was specialized;
- **general classifier agents** are trained to diverse categories of images, and they receive a constant weight parameter;
- The **coordinator agent** (or coordination agent) responsible for an specific file asks for specialized and general agents to classify this file; when there is not enough available agents, this queen instantiates new workers of both types and attributes the classification for them;

- When the attributed workers realize the classification, the queen uses a fuzzy inference system to combine the results of these workers according to their fuzzy weights and finds her classification result.

When more than one coordination agent come to divergent results, they start a negotiation process in order to find the final classification result.

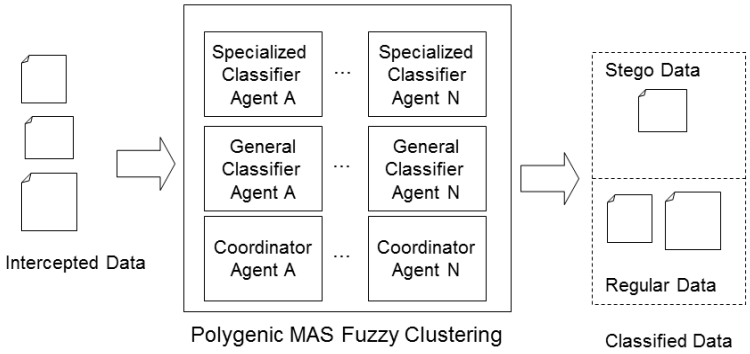


Fig. 2. Polygenic MAS Fuzzy Clustering Steganalysis Architecture.

The specialized and the general classifier agents represent different specialized workers in this metaphor. The communication between the agents is realized through a message board.

The accuracy rate for the general classifier agents and the specialized classifier agents can be fuzzyficated as shown in the graphic bellow. Where L, M and H means Low, Medium and High accurate rates, respectively (Figure 3.a). In order to linearize the classification problem, the classification process will give a probability a given sample is or not a stego object according to the classification (Figure 3.b). The inference method used is a simple Mamdani FIS, and the Figure 4 simplifies its mechanisms.

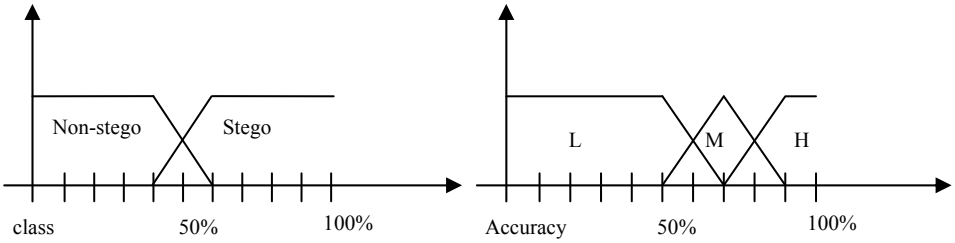


Fig. 3. Fuzzyfication. a) accuracy rates; b) probability of a classified sample being stego object.

3.2 Use case view

Figure 5 presents the use case diagram of the approach. The use cases presented are: monitoring of files, negotiation of final result, coordination of classification, attributing/instantiating agents, and classification; and these actions are realized as described next. The actors that are present at the use case of the system are the monitors (monitor agent),

queens (coordinator agents), and laborers (general and specialized classifier agents). The role of **monitor** can be performed by simple users of the system that submit a set of files to be monitored by the system agents to work on them. Alternatively, one can program monitor agents to perform searchers or sniffs in the internet in order to collect and analyze data.

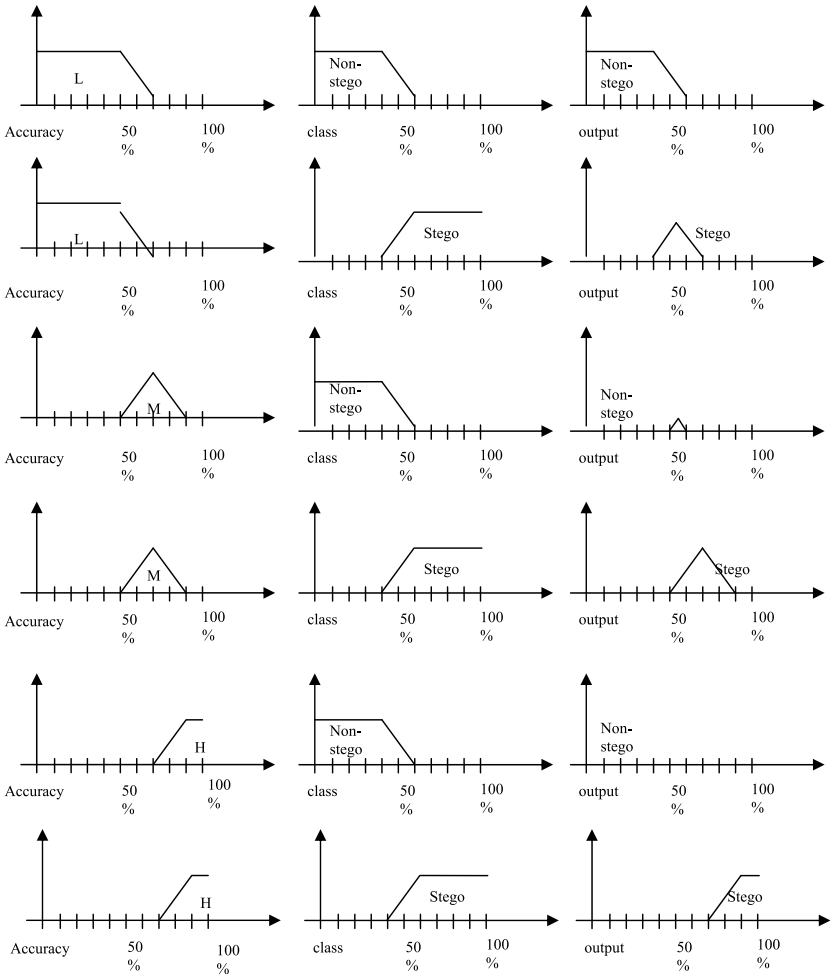


Fig. 4. Fuzzy inference system for classification of stego images (simplification).

After the monitor agent (or user) perform the monitoring of files, random coordination agents are attributed that file and individually start coordination the global classification process. This process, in his turn, needs the attribution or instantiation of general and specialized classifier agents to the task of classifying the monitored file. When a instantiation is needed, the agents are trained and tested in order to receive a weight corresponding to their adequacy in the classification process. So, the roles of general and specialized classification agents are only responsible to classify the data and send this result to the

coordination agent, which will use it fuzzy clustering inference algorithm to define the classification result. After the coordination agents find their classification results, if divergent, they negotiate to find a final answer.

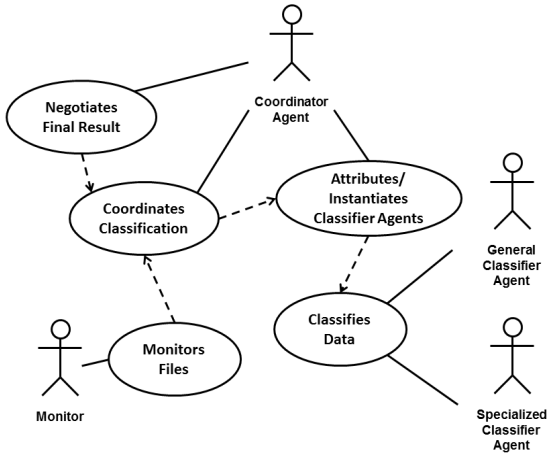


Fig. 5. Use Case Diagram of Polygenic MAS Fuzzy Clustering Steganalysis.

All the inference rules of the negotiation protocol follow:

1. IF TC=OC THEN accord
2. IF TAAR=High and OAAR=High and TC!=OC and TAAR>OAAR THEN TAAR=TAAR+0.01
3. IF TAAR=High and OAAR=High and TC!=OC and TAAR=OAAR and TAFN < OAFN THEN TAAR=TAAR+0.01
4. IF TAAR=High and OAAR=High and TC!=OC and TAAR=OAAR and TAFN >= OAFN THEN TAAR=TAAR-0.01
5. IF TAAR=High and OAAR=High and TC!=OC and TAAR<OAAR THEN TAAR=TAAR-0.01
6. IF TAAR=High and OAAR=Medium and TC!=OC THEN class=TC
7. IF TAAR=High and OAAR=Low and TC!=OC THEN class=TC
8. IF TAAR=Medium and OAAR=High and TC!=OC THEN class=OC
9. IF TAAR=Medium and OAAR=Medium and TC!=OC and TAAR>OAAR THEN TAAR=TAAR+0.01
10. IF TAAR=Medium and OAAR=Medium and TC!=OC and TAAR=OAAR and TAFN < OAFN THEN TAAR=TAAR+0.01
11. IF TAAR=Medium and OAAR=Medium and TC!=OC and TAAR=OAAR and TAFN >= OAFN THEN TAAR=TAAR-0.01
12. IF TAAR=Medium and OAAR=Medium and TC!=OC and TAAR<OAAR THEN TAAR=TAAR-0.01
13. IF TAAR=Medium and OAAR=Low and TC!=OC THEN class=TC
14. IF TAAR=Low and OAAR=High and TC!=OC THEN class=OC
15. IF TAAR=Low and OAAR=Medium and TC!=OC THEN class=OC
16. IF TAAR=Low and OAAR=Low and TC!=OC and TAAR>OAAR THEN TAAR=TAAR+0.01

17. IF TAAR=Low and OAAR=Low and TC!=OC and TAAR=OAAR and TAFN < OAFN THEN TAAR=TAAR+0.01
18. IF TAAR=Low and OAAR=Low and TC!=OC and TAAR=OAAR and TAFN >= OAFN THEN TAAR=TAAR-0.01
19. IF TAAR=Low and OAAR=Low and TC!=OC and TAAR<OAAR THEN TAAR=TAAR-0.01

where,

TAAR - this agent accuracy rate
OAAR - other agent accuracy rate
TC - class according to this agent
OC - class according to the other agent
high - accuracy rate is greater than or equal 0.8
medium - accuracy rate is greater than or equal 0.6 and lesser then 0.8
low - accuracy rate is lesser than 0.6
class - the new result of this agent
accord - finish the negotiation process
TAFN - this agent false negative rate
OAFN - other agent false negative rate

3.3 Description of steganalysis

For the approach described here, we assume that suspicious data is already intercepted and submitted to this approach in order to verify if an object is a stego object or a regular file.

3.3.1 Classifier method

The classifier method used in this work is the Polygenic MAS Fuzzy Clustering. In a MAS architecture, coordination agents use fuzzy clustering inference to group the classification result of specific and general classification agents. Also, negotiation is performed between coordination agents in order to decide the better result.

The classifiers of two specific classification agents are divergent, both because each can be trained for a different type of image, and because each receive a different training subset of the data samples. This last reason also applies to describe the difference between two general classification agents.

The machine learning algorithm chosen for the internal classifier of each agent is a Decision Tree. But different machine learning algorithms would be applied to this architecture in order to search for a more robust classification.

3.3.2 Features calibration

The features that describe each instance or sample are: the four statistical momentums (mean, variance, asymmetry and kurtosis) for both the RGB and the HSBr matrixes, the image category (smooth, regular, complex, noisy), and the name of the class that sample describes (stego/non-stego).

This configuration of features was settled after some tryouts with bigger feature lists, which included:

1. ***region-centroid-col****: central column in a 3x3 pixel region;*
2. ***region-centroid-row****: central line in a 3x3 pixel region;*
3. ***region-pixel-count****: total of pixels in a 3x3 region = 9.*
4. ***short-line-density-5****: counts how many lines with low contrast and size lesser or equal to 5 passes through the region;*
5. ***short-line-density-2****: counts how many lines with high contrast and size greater or equal to 5 passes through the region;*
6. ***vedge-mean****: vertical edge mean;*
7. ***vegde-sd****: vertical edge standard-deviation;*
8. ***hedge-mean****: horizontal edge mean;*
9. ***hedge-sd****: horizontal edge standard-deviation;*
10. ***intensity-mean****: (R + G + B)/3 in a region;*
11. ***rawred-mean****: red mean in a region;*
12. ***rawblue-mean****: blue mean in a region;*
13. ***rawgreen-mean****: green mena in a region;*
14. ***exred-mean****: additional red mean: (2R - (G + B));*
15. ***exblue-mean****: additional blue mean: (2B - (G + R));*
16. ***exgreen-mean****: additional green mean: (2G - (R + B));*
17. ***value-mean****: non-linear 3D transformation mean;*
18. ***saturation-mean****: saturation mean in the 3D transform;*
19. ***hue-mean****: hue mean in the 3D transform;*

The features above were used based in literature review, where we choose to operate in special domain instead of transform domain, by empirical experimentation. Though, we observed that the efficiency of the machine learning methods did not decrease by eliminating many of the features above, so they were excluded from the final features list.

3.3.3 Data samples

To create the dataset, we utilized 300 images of landscapes, interiors, animals, buildings, people and food. According to the graphical complexity of each image, they were categorized as smooth, regular, complex or noisy. A random half of these images were kept unmodified, while the other half received hidden data corresponding up to 10% of the carry images size, what represents 20% of the maximum payload a carry image may cover (which is 50% of the total size of the carry image). The average size of the cover images is 800 x 600 pixels. And the stego objects here were created with the steganographic method JPHide/JPSeek (Lathan 2006).

Then, this dataset was once again divided. A random 80% of all the images were separated to compose the training set, and the 20% left composed the testing set.

3.3.4 Training and testing

At runtime, when a specialized classifier agent is instantiated, it receives a random subset from the training set. This subset is selected from all the samples that correspond to one single of the four image categories used in this work (smooth, regular, complex, noisy). Three other specialized classifier agents are created to the other categories. The subset for each of these agents is 20% the size of the training set samples. Similarly, when a general

classifier agent is instantiated, it receives a random subset from the training dataset that corresponds to 40% of the total size of the training dataset.

Then, these agents train theirs classifiers and test their performances with 20% of the training dataset (the testing dataset is for testing the approach as a whole). The resulting accuracy rate will be informed to the coordinator agent, generating a weight to that agent decision, and influencing the fuzzy inference mechanism.

The size of these datasets was limited to this small percent in order to produce different agents, with different performances, that will be combined by the polygenic MAS fuzzy clustering approach.

4. Experiments and results

We have developed experiments and tests following the planning experimentation setup discussed by Cobb (1997). Basically, this methodology is resumed in *what measures to take, under what conditions and which material to process in the testes*. The answers are the measures given by the MAS about the classification: correctness rate, false positive and false negative. A set of training data is presented to the system, which is randomly distributed into other data sets for training the classifier agents. The test data is then distributed between the coordinator agents in order for these to coordinate classification activities of the general classifier agents and the specific classifier agents. Finally, the cited rates are obtained and analyzed.

As result of these experiments, the system presents a rate for correct detection of 89,37%, with false positive of 10,63% and false negative of 10,54%.

For JPHide and Seek, Liu (2008) present accuracy rate of 0.8% with OAASVM, and 56% with Adaboost. It is important to say that their dataset is different from the one used here, although we may say that our experiment presented a considerate accuracy rate.

Bakhshandeh (2009) presented accuracy rates from 68,75% to 94,67%, but none of the steganography methods used in their experiment was the same used in our work or in Lius´.

Also, we trained a Decision tree without the Polygenic MAS Fuzzy Clustering Steganalysis, using the entire dataset. And the results were an accuracy rate of 72,45%, false positives of 27,23% and false negatives of 26,92%.

These results show that the Polygenic MAS Fuzzy Clustering Steganalysis approach increased the performance of a machine learning steganalysis.

5. Conclusion

We have proposed a useful technique to detect images that possibly carry encrypted data on its contents. We use a methodology based on polygenic bees (a model based on community) where several agents interact between them. Our model combines this multi-agent system system with fuzzy logic in order to decide whether a digital media object has hidden information, coming up with a decision at the end of processed interactions. In comparison to the rates of correctness of other techniques found in the literature (between 70% to 90%) our rates of about 89% indicates that the proposed approach based on fuzzy logic is a good choice in this direction, being efficient in this task, experimentally comproved.

As future work we intend to to improve this paradigm once the use of MAS can be extended to other learning techniques besides fuzzy logic. A comparison between several techniques will be performed and a possible solution combining two or several of them will also be tried in order to achieve even better performance. For example, we believe that the use of decision trees combined to our fuzzy approach depicgted here can be used hopefully to get better results. So a possible future direction for our work is to test this approach with other techniques and also to use other medias as video, text, and audio, not being addressed in this work. Finallyanother possibility is to develop a more complete system including techniques for extracting the hidden information.

6. References

Acampora 2011 Acampora, G.; Cadenas, J.M.; Loia, V.; Ballester, E.M.;A Multi-Agent Memetic System for Human-Based Knowledge Selection, in IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol.41, no.5, pp.946-960, Sept. 2011.

Amirkhani 2011 Amirkhani, Hossein; Rahmati, Mohammad: New framework for using image contents in blind steganalysis systems. in Journal of Electronic Imaging, Vol.20, Iss.1, pp.013016, 2011, ISSN: 10179909

Aponte 2003 Aponte, Olga Inés Cepeda: Poliginia e monoginia em Melipona bicolor (Apidae, Meliponini): do coletivo para o individual. DSc Thesis in Biosciences, Universidade de São Paulo, 2003.

Arroyo 2011 Arroyo, A.; Serradilla, F.; Calvo, O. Adaptive fuzzy knowledge-based multi-agent systems on virtual environments. in Expert Systems - Special Issue: New Perspectives on the Application of Expert Systems, v. 28 (4), pp 339–352, September 2011.

Artz 2001 ARTZ, Donovan. Digital Steganography: Hiding Data within Data. IEEE Internet Computings, v.5, n.3, mai.-jun. 2001.

Bakhshandeh 2009 Bakhshandeh, Soodeh; Jamjah, Javad Ravan; Azami, Bahram Zahir: Blind Image Steganalysis Based on Local Information and Human Visual System. in Signal Processing, Image Processing and Pattern Recognition. V. 61: 201-208. SPRINGER-VERLAG, BERLIN, 2009.

Carpenter 1992 Carpenter, G.A.; Grossberg, S.; Markuzon, N.; Reynolds, J.H.; Rosen, D.B.: Fuzzy Artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps in IEEE Transactions on Neural Networks, vol.3, no.5, pp.698-713, Sep 1992.doi: 10.1109/72.159059

Chandramouli 2004 Chandramouli, R.; Subbalakshmi, K.P: Current Trends in Steganalysis: A Critical Survey, Invited session on Multimedia Security, The Eighth International Conference on Control, Automation, Robotics and Vision, ICARCV 2004, December 2004. (invited paper)

Fridrich 2002 Fridrich, J., Goljan, M.: Practical steganalysis of digital images - state of the art. In: Proc. of SPIE Photonics West. Volume 4675. San Jose, California, USA – 2002.

Fridrich 2004 Fridrich, J.; Goljan, M.: On estimation of message length in LSB steganography in spatial domain. Security, Steganography, and Watermarking of Multimedia Contents , 2004.

Goodarzi 2011 Goodarzi, Mohammad; Radmand, Ashkan; Nazemi, Eslam: An Optimized Solution for Multi-agent Coordination Using Integrated GA-Fuzzy Approach in Rescue Simulation Environment, in Advances in Practical Multi-Agent Systems

Studies in Computational Intelligence, 2011, V. 325, pp 377-388, DOI: 10.1007/978-3-642-16098-1_23

Hagras 2010 Hagras, H.; Ramadan, R.; Nawito, M.; Gabr, H.; Zaher, M.; Fahmy, H.: A fuzzy based hierarchical coordination and control system for a robotic agent team in the robot Hockey competition, in IEEE International Conference on Fuzzy Systems (FUZZ), pp.1-8, July 2010. ISSN: 1098-7584.

Jang 1993 Jang, JSR: Anfis - Adaptive-Network-Based Fuzzy Inference System. in IEEE Transactions on Systems Man And Cybernetics, 23(3), pp665-685. May-Jun 1993. DOI: 10.1109/21.256541

Katzenbeisser 2000 Katzenbeisser, Stefan; Petitcolas, Fabien A. P. Information Hiding Techniques for Steganography and Digital Watermarking. Boston: Artech House, 2000.

Kofler 2003 Kofler, R. Krimmer, R. Prosser, A.: Electronic Voting: Algorithmic and implementation Issue. System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on , 6-9 Jan. 2003.

Lin 2002 Lin, Chun-Fu; Wang, Sheng-De.: Fuzzy support vector machines in IEEE Transactions on Neural Networks, vol.13, no.2, pp.464-471, Mar 2002.

Liu 2008 liu, qingzhong; sung, Andrew H.: Detect Information-Hiding Type and Length in JPEG Images by Using Neuro-fuzzy Inference Systems, CISP, vol. 5, pp.692-696, 2008 Congress on Image and Signal Processing, Vol. 5, 2008.

Liu 2008b Liu, Q., Sung, A.H., Chen, Z., Xu, J.: Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images. Pattern Recognition, 41 (1), pp. 56-66. 2008.

López-Ortega 2011 López-Ortega, Omar; Rosales, Marco-Antonio: An agent-oriented decision support system combining fuzzy clustering and the AHP, in Expert Systems with Applications, Vol. 38 (7), July 2011, pp 8275-8284, ISSN 0957-4174.

Luciano 2003 Luciano, E. M.; Testa, M. G.; Freitas, H. : As tendências em comércio eletrônico com base em recentes congressos. XXXVIII CLADEA, Lima/Peru, 2003.

Macedo 2001 Macedo, A. P. Cunha: Metodologias de Negociação em Sistemas Multi-Agentes para Empresas Virtuais. Doctors Thesis, Faculdade de Engenharia, Universidade do Porto, 2001.

Macq 1995 Macq, B. M.; Quisquater , J-J. Cryptology for digital TV broadcasting. Proceedings of the IEEE , 1995.

Moskowitz 2002 MOSKOWITZ, Ira S.; CHANG, Liwu; NEWMAN, Richard E. Capacity is the wrong paradigm. In *Proceedings of the 2002 workshop on New security paradigms* (NSPW '02). ACM, New York, NY, USA, 114-126. DOI=10.1145/844102.844124 http://doi.acm.org/10.1145/844102.844124

Pavlidis 1982 Pavlidis, T. Algorithms for graphics and image processing, Springer, Berlin, 1982.

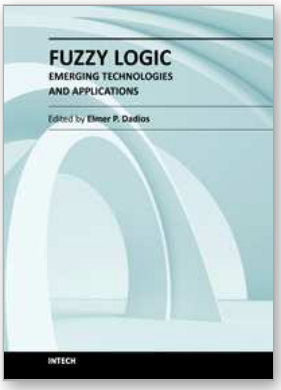
Rezende 2003 Rezende, Solange Oliveira. Sistemas Inteligentes: Fundamentos e Aplicações. Manole Editora. 2003. 525p

Russel 1995 Rusell, Stuart J.; Norving, Peter. Artificial Inteligence: A Modern Approach. Prentice-Hall Series in Artificial Inteligence, 1995.

Sanches 2004 Sanches, M. K.; Geromini, M. R.; Aprendizado de Máquina: Relatório Técnico. Instituto de Ciências Matemáticas e Computação, Universidade de São Paulo, 2004.

Wooldridge 2001 Woolridge, Michael J., Introduction to Multiagent Systems, John Wiley & Sons, Inc., New York, NY, 2001.

Zadeh 1965 Zadeh, L. A.: Fuzzy sets. Information and Control, 8(3), pp. 338-353, 1965.



Fuzzy Logic - Emerging Technologies and Applications
Edited by Prof. Elmer Dadios

ISBN 978-953-51-0337-0
Hard cover, 348 pages
Publisher InTech
Published online 16, March, 2012
Published in print edition March, 2012

The capability of Fuzzy Logic in the development of emerging technologies is introduced in this book. The book consists of sixteen chapters showing various applications in the field of Bioinformatics, Health, Security, Communications, Transportations, Financial Management, Energy and Environment Systems. This book is a major reference source for all those concerned with applied intelligent systems. The intended readers are researchers, engineers, medical practitioners, and graduate students interested in fuzzy logic systems.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821