Integrated Method for Detection of Unknown Steganographic Content

Magdalena Pejas

Abstract**—**This article concerns the presentation of an integrated method for detection of steganographic content embedded by new unknown programs. The method is based on data mining and aggregated hypothesis testing. The article contains the theoretical basics used to deploy the proposed detection system and the description of improvement proposed for the basic system idea. Further main results of experiments and implementation details are collected and described. Finally example results of the tests are presented.

Keywords**—**Steganography, steganalysis, data embedding, data mining, feature extraction, knowledge base, system learning, hypothesis testing, error estimation, black box program, file structure.

I. Introduction

STEGANOGRAPHY is the art of hiding data in another data. The problem described in the article concerns the question: "Is it possible to detect any steganographic content in digital images with a reliable detection level in the case, when we don't know the algorithms of steganographic programs which can be potentially used for data embedding?". The answer is YES, when we define the field of the problem and make several helpful assumptions.

II. Theoretical Basics

A. Basic Assumptions and Problem Field Constraints

In short it can be asserted that the experiments are constrained to the following situation:

1) We have a set of digital images created with any digital camera and before the experiments they haven't been processed by any program.

2) We have a fixed set of steganographic programs, which can be run in the tests, but we don't know how they work in details.

3) We learn the system to detect data embedded by programs with the defined already set.

4) We develop the proper detection system and learn it with the use of Bayesian criterion customized for steganalysis.

B. The Bayesian Criterion for Object Classification

Book [1] contains the theory of Bayessian criterion used for signal detection. This theory can be easily applied for the detection of steganographic data in digital objects.

In the simplest binary system of signaling there are two hypotheses: 0-hypothesis denoted as H_0 and 1-hypothesis denoted as H_1. The probability of the 0-hypothesis is denoted as $P(H_0)$ and for the 1-hypothesis as $P(H_1)$. The threshold used for Bayesian classification should fulfill the following condition:

$$\lambda_{th} = \frac{P(H_0)}{1-P(H_0)}\frac{c_1}{c_2}, \qquad (1)$$

where c_1 and c_2 refer to the costs of the errors of the first and the second type, respectively. The ratio (1) refers to the value x_0, which is the classification threshold.

The Fig. 1 represents two probability density functions referring to each of the hypotheses.

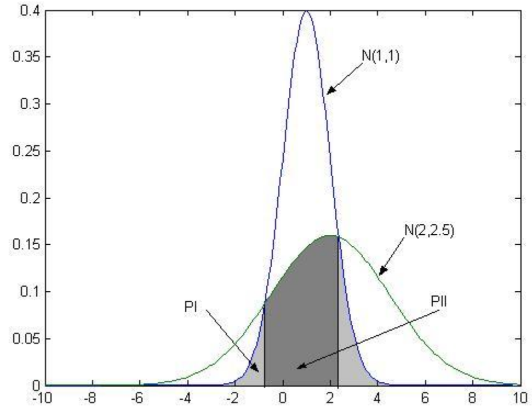

Fig. 1 The probability density functions and classification errors

The gray spaces marked in the figure represent the classification errors of the first and the second type. The error of the first type takes place when we make decision D_1 while the hypothesis H_0 is true and vice versa.

The cost of all possible errors in generalized multi hypothesis case is defined in (2):

$$C_{total} = \sum_{i=0}^{N}\sum_{j=0}^{N} P(D_i \mid H_j)P(H_j)C_{i\,j}\,. \qquad (2)$$

Manuscript received September 30, 2006. This work was supported in part by The Warsaw Scientific Society.

M. P. is with the Warsaw University of Technology, The Faculty of Electronics and Information Technology, Warsaw, Poland (phone: +48-0-888-552-778; e-mail: magdap7@ gazeta.pl).

C. *Applying of Bayesian Classificator for Steganalysis*

In the filed of steganalysis the 1-hypothesis is combined with the situation, that a given object contains embedded data. The 0-hypothesis refers to the clean object.

The error of the first type refers to the situation that we miss an object infected by steganographic content. By analogy – the error of the second type is the false alarm error.

D. *The Set of Representative Tests*

In steganalysis there are many various tests used for feature extraction and statistics measures. In [2-10] there are described main popular steganalytic tests used in worldwide university experiments.

The tests can be grouped to the following families:

1) Histogram distortion tests,
2) Correlation and prediction tests,
3) Binary and pattern similarity tests,
4) Data format compatibility tests,
5) Noise and entropy level measures,
6) Pseudo-randomness tests,
7) Image and Audio quality measures.

In this article I concentrate on a given subset of correlation tests. Apart form that I present two additional groups of developed tests: signature tests and file structure anomaly tests.

As we see later, grouping of tests is useful in applying the aggregation of hypothesis testing.

III. The Proposed Improvements

In this chapter I will present several improvements added to the main conception of the Bayesian based steganalytic system. These are listed as follows:

- Grouping tests into families according to their simplicity and detection level,
- Adding new tests for better detection of fingerprints left by specific programs,
- Proposing a heuristics in choosing next tests in the testing loop,
- Adding “post-testing” phase to refine the value of the classification threshold even in the testing phase.

A. *Selection and Grouping of Representative Tests*

In this article I consider three classes of tests:

- fingerprint tests,
- local structure anomaly tests,
- binary correlation tests.

The first type of testing concerns checking, if a given region of a digital object is equal to a given static or dynamic flexible pattern defined by a given syntax rule, for example:

$[101[0|1]^3 111]$.

Tests of the second type check more generalized situations, where a region of data with odd exists or abnormal random density function.

Expression (3) contains an interesting generalization of correlation tests. From this definition we can conclude many kinds of filters and smoothness tests.

$$T(O) = \frac{1}{N(O)} \sum_{i=2}^{N(O)} (O_i \bullet O_{i-1}) \,, \tag{3}$$

where O is a given tested object, N is the number of units (e.g. pixels) in the object and (4) defines several mathematical conditions and operations taken to the consideration:

$$\bullet \in \{<, >, \leq, \geq, =, +, -, \oplus, \otimes, \square \} \,. \tag{4}$$

Expressions (5-8) present the mathematical description of basic tests used in the experiments:

$$\frac{1}{N} \sum_{i=2}^{N} | X_i - X_{i-1} | \tag{5}$$

$$\frac{1}{MN} \sum_{i=2,\, j=1}^{M \cdot N} | X_{i,j} - X_{i-1,j} | \tag{6}$$

$$\frac{1}{MN} \sum_{i=2,\, j=2}^{(M-1)\cdot(N-1)} | 2 \cdot X_{i,j} - X_{i,j-1} - X_{i,j+1} | \tag{7}$$

$$\frac{1}{MN} \sum_{i=2,\, j=2}^{(M-1)\cdot(N-1)} | 2 \cdot X_{i,j} - X_{i-1,j} - X_{i-1,j} | \tag{8}$$

$$\frac{1}{MN} \sum_{i=2,\, j=2}^{(M-1)\cdot(N-1)} | 4 \cdot X_{i,j} - X_{i,j-1} - X_{i,j+1} - X_{i-1,j} - X_{i+1,j} | \tag{9}$$

The Table I contains the description of the given smoothness tests.

TABLE I
Most Representative Correlation Tests

No	Ref (n)	Description
1	(5)	Smoothness of a vector
2	(6)	Smoothness of a matrix
3	(7)	Horizontal neighbors contrast
4	(8)	Vertical neighbors contrast
5	(9)	Horizontal and Vertical neighbors contrast

B. *Developing New Tests*

As mentioned earlier, two additional classes of tested were introduced. First – fingerprint tests and the second – local structure anomaly tests.

The threshold values have form dependent on the class of tests:

- static string, e.g. [10101010],
- description of a pattern, e.g. [10[1 | 0] [1 | 0]01],
- the pair of min and max threshold value, e.g. (1.26, 1.32).

Similarly the testing conditions have different types of expressions.

C. *Refining Bayes Classificator*

The quality of a given test depends on the cost and the probability of the first and the second type of errors. The error of the first type can be estimated according the following formula (10):

$$P(E_{01}) = P(D_0 \mid H_1) = \int_{R_0} P_1(x)dx \qquad (10)$$

and similarly for the false alarm error:

$$P(E_{10}) = P(D_1 \mid H_0) = \int_{R_1} P_0(x)dx\,. \qquad (11)$$

In theory and in practice the smoothness tests can be optimized by 3-dim integration. The false negative error is much less than in case of 1-dim for every color independently. This is because no information is ignored in the case of 3-dim integration. Expressions (12-13) show the mathematical formula for this approach.

$$P(E_{10}) = \int_{x_r \in R_1} \int_{x_g \in G_1} \int_{x_b \in B_1} P_0(x)dx_r dx_g dx_b\,, x \in R^3 \qquad (12)$$

$$P(E_{01}) = \int_{x_r \in R_0} \int_{x_g \in G_0} \int_{x_b \in B_0} P_1(x)dx_r dx_g dx_b\,, x \in R^3 \qquad (13)$$

The Table II lists the values of errors of the all of 4 types. From the results we can observe, that for three-dimensional integration the errors become not significant.

TABLE II
ERROR LEVELS FOR 1, 2 AND 3 DIMENSIONAL TESTS

		P_{10}	P_{00}	P_{11}	P_{01}
$P(X), X \in R^1$	T_{sm1}	0.4812	0.5187	0.7812	0.2188
	T_{sm2}	0.3750	0.6250	0.8000	0.2000
	T_{sm3}	0.3750	0.6250	0.8187	0.1813
	T_{sm4}	0.3562	0.6437	0.8562	0.1438
$P(\overrightarrow{\overrightarrow{X}}), \overrightarrow{\overrightarrow{X}} \in R^2$	T_{sm1}	0.2438	0.7562	0.9500	0.0500
	T_{sm2}	0.0125	0.9875	0.9937	0.0063
	T_{sm3}	0.0313	0.9687	1.0000	0
	T_{sm4}	0.0063	0.9937	1.0000	0
$P(\overrightarrow{\overrightarrow{\overrightarrow{X}}}), \overrightarrow{\overrightarrow{\overrightarrow{X}}} \in R^3$	T_{sm1}	0.2937	0.7062	0.8312	0.1687
	T_{sm2}	0.0313	0.9687	0.9937	0.0063
	T_{sm4}	0.0500	0.9500	1.0000	0
	T_{sm1}	0.0375	0.9625	1.0000	0

D. Adding Heuristics

As mentioned earlier, grouping of tests can be helpful for choosing proper heuristics for choosing the order of tests while performing general test.

$$Ord(t) \square \frac{P(H_1)}{O(t)} P_{err}\,. \qquad (14)$$

The priority of choosing a given test in each test family is proportional to the frequency of data embedding with a given steganographic program related to the test and the detection level estimated for the test. The groups of test are sequenced accordingly to the complexity of the algorithm of a given test family.

E. Adding Post Training Phase

Performing the post-training updates of knowledge base is possible if the following condition is fulfilled:

$$P_e : P_{10} <= \delta_{10} \wedge P_{01} <= \delta_{01}, where\ \delta_{10} << 1 \wedge \delta_{01} << 1 \qquad (15)$$

The probability of the 1 and 0- hypothesis is updated accordingly to the result of the recent test.

$$P'(H_0) = \begin{cases} \dfrac{N \cdot P(H_0)+1}{N+1}, & for\ D_0 \\ \dfrac{N \cdot P(H_0)}{N+1}, & for\ D_1 \end{cases}, \qquad (16)$$

where D_0 corresponds to the 0-decision and D_1 - to the 1-decision. The symbol N denotes the number of already tested objects and it is also updated:

$N' = N + 1$.

Then we can calculate the current probability of errors of the first and the second type denoted as P_{err}. It is dependent on the current decision threshold λ_{thr} and the probability density function P_{dist} accordingly to (10) and (11). The decision threshold depends on the costs of errors C_{err} and the probability of hypotheses and is calculated from (1). The pseudo-code is shown in the Listing 1.

```
$in : \{C_{err}, P_{hyp}, P_{dist}, P_{err}, D\}$
    $P'_{hyp}(D, P_{hyp})$
    $\lambda'_{thr}(C_{err}, P'_{hyp}, P_{dist})$
    $P'_{err}(P_{dist}, \lambda'_{thr})$
$out : \{P'_{hyp}, P'_{err}\}$
```

Listing 1 A single iteration of the post-training phase

The sense of applying the post training phase comes from the observation that in every new test, when condition (15) is fulfilled, the probabilities of errors of both types tend to lower.

IV. THE IDEA OF THE SYSTEM

The system consists of the learning part and the testing part. In the learning part the knowledge base is filled with valid threshold values. Each value corresponds to a given steganographic program.

In the testing part each of new-tested objects has its features extracted and compared with the appropriate threshold values. The object is classified either as “infected” or “clean with high level of probability”.

In the developed version of the system the decision parameters are updated also during the testing part.

More about the basics of the conception of the system and specific tests can be read in my earlier papers [11-12].

A. *The Learning Part*

In the learning part we have two sets of training objects: provable clean objects and objects containing data embedded with a fixed steganographic program. Several tests are performed on them and then the initial decision values are estimated and saved in the knowledge base.

B. *Knowledge Base*

The knowledge base consists of the test descriptors. Table III depicts the construction of the structure describing each test. Example descriptors are presented in the subchapter concerning the final results of the experiments.

TABLE III
TEST DESCRIPTOR

No	Structure field	Description
1	Class	The class a given test belongs to
2	Name (program name)	The program to which the test refers
3	Description	The region of image which is tested
4	Decision value	The value which is tested in the test
5	Threshold value	The threshold for classification
6	Condition description	The condition defining the test
7	Implementation (Matlab)	Implementation of the test
8	$P = \{P_{10}, P_{11}, P_{00}, P_{01}\}$	The detection quality of the test
9	$O(f(n))$	Estimated complexity of the test

C. *The Testing Part*

All tests are treated as objects of a given class. The tests belonging to one class have similar structure, testing conditions and algorithms. The tests from one class differ from each other only in parameters such as decision values and threshold values.

In the testing part the tested object is checked with sequence of tests from each test class. Test classes with lower complexity are taken first. In each family the order of tests is dependent on the frequency of the positive decisions. If any test returns positive decision, the process is stopped.

D. *Aggregation of Hypotheses*

The Fig. 2 depicts the idea of hypotheses tests aggregation.

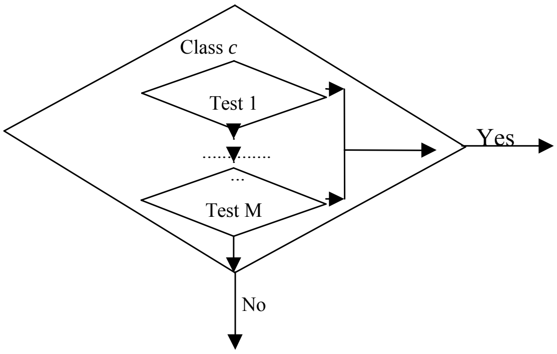

Fig. 2 The idea of test aggregation

The aggregation of hypotheses is defined in the expression (17). The sum of probabilities of the positive hypotheses in one class is equal to the probability of the positive hypothesis assigned to the class to which these tests belong.

$$P(H^{cl}) = \sum_{t=1}^{Nt(cl)} P(H_t^{cl}). \tag{17}$$

where given test *t* belongs to a fixed class *cl*. *Nt* is the number of tests belonging to the class. It works if we can make an assumption that every embedding process removes the results of the previous data embedding. In conclusion we can write:

$$P(H_0) + \sum_{cl=1}^{Nc} P(H^{cl}) = 1 \tag{18}$$

This works in ideal model of steganalytic system, but in practice in many cases we can observe an interference of effects caused by data embedding. This means we can obtain positive results from more than one test.

V. IMPLEMENTATION AND RESULTS

In this subchapter I present some implementation details and example descriptors of single tests one from each class. Every test is defined and has its quality estimated.

I take into consideration three classes:

- signature test,
- local structure anomaly test,
- smoothness tests.

We can observe that the descriptors of the tests from the same class, group, and family look similar. This grouping helps to organize and manage the set of all of the tests. It also makes possible to better optimize the testing process.

A. *Signature Tests*

The descriptor of an example signature test is presented in the Table IV.

TABLE IV
SIGNATURE TEST DESCRIPTOR

Signatures
Courier 1.0
Finds a fixed string in the second row of the image
lsb(m)=bitand(1,m(2,:,c)),c = 1..3
$\mathrm{sgn}_1 = [10]^*$
$\forall_{x\in<1,width(m)>,c\in\{r,g,b\}}$ lsb(m(2,x,c))== $\begin{cases} 0, x \in nP \\ 1, x \in P \end{cases}$

s1(c)=sum(lsb(2,1:2:Y,c)==0);
s2(c)=sum(lsb(2,2:2:Y,c)==1);
test(c)=(abs(s1+s2-Y)<=2);

$P \cong \{0,1,1,0\}$
$O(width(image))$

B. *Structure Anomaly Tests*

The descriptor of an example anomaly test is presented in the Table V.

TABLE V
STRUCTURE ANOMALY TEST DESCRIPTOR

Anomalies
WbStego 4.2
Finds a string of blank characters in the end of lines.
d=dataLoad(filename);L=length(d);d(L-7:L)
<0,32>
$\underset{x}{\exists}\ \underset{s\in<x,length(line)}{\forall}\ s\in<0,32>$

thr=[0,32];
s=sum(d(L-7:L)>=thr(1) & d(L-7:L)<=thr(2))
test=(s==8)

$P_{10}(n)=P(\underset{i=1..n}{\forall} b_i \in <0,32>)=\left(\frac{32}{256}\right)^n \cong (\frac{1}{8})^n$
$O(n)$

C. Correlation and Smoothness Tests

The descriptor of an example smoothness tests is presented in the Table VI.

TABLE VI
SMOOTHNESS TEST DESCRIPTOR

Smoothness / correlation tests
a) Cloak 0.5
Performs smoothness tests and compares with the threshold value.
{sm3,sm4}
λ
{sm3,sm4}
sm_1=m(2:x,2:y)-m(1:x-1,1:y-1); sm_2=2*m(2:x-1,2:y-1)-m(1:x-2,2:y-1)-m(3:x,2:y-1); sm_3=2*m(2:x-1,2:y-1)-m(2:x-1,1:y-2)-m(2:x-1,3:y); sm_4=4*m(2:x-1,2:y-1)-m(1:x-2,2:y-1)-m(3:x,2:y-1)-m(2:x-1,1:y-2)-m(2:x-1,3:y);
$P_{T_{sm2}}=\{0.0313,0.9687,0.9937,0.0063\}$ $P_{T_{sm3}}=\{0.0500,0.9500,1.0000\ ,0\}$ $P_{T_{sm4}}=\{0.0375\ ,0.9625,1.0000,0\}$
$O(n^2)$

D. Implementation

The Table VII includes the Matlab implementation of the basic formulas given above in this paper.

TABLE VII
IMPLEMENTATION OF CALCULATIONS

Formal description	Matlab Implementaion
$P(x)$	$\overrightarrow{h_i}=hist(\overrightarrow{X_i})/length(\overrightarrow{X_i});$
$\lambda_{thr}=\frac{P(H_0)}{1-P(H_0)}\frac{c_1}{c_2}$	$\overrightarrow{ind_i}=find(\overrightarrow{h_j}<\overrightarrow{h_i}\cdot\lambda_{thr});$
$P(E_{ij})=\int_{R_i} P_j(x)dx$	$\overrightarrow{ind_i}=find(\overrightarrow{h_j}<\overrightarrow{h_i});$ $P_{ij}=sum(\overrightarrow{h_j}(\overrightarrow{ind_i}));$
$P(E_{10})=\int_{x\in X_1} P_0(x)dx$	$\overrightarrow{ind_1}=find(\overrightarrow{h_0}<\overrightarrow{h_1});$ $P_{10}=sum(\overrightarrow{h_0}(\overrightarrow{ind_1}));$
$P(E_{01})=\int_{x\in X_0} P_1(x)dx$	$\overrightarrow{ind_0}=find(\overrightarrow{h_0}>\overrightarrow{h_1});$ $P_{01}=sum(\overrightarrow{h_1}(\overrightarrow{ind_0}));$
$P(E_{01})=\int_{x_r\in R_0}\int_{x_g\in G_0}\int_{x_b\in B_0} P_1(x)dx$	$\overrightarrow{ind_0}=find(\overrightarrow{\overrightarrow{\overrightarrow{h_0}}}>\overrightarrow{\overrightarrow{\overrightarrow{\overrightarrow{h_1}}}});$ $P_{01}=sum(\overrightarrow{\overrightarrow{\overrightarrow{h_1}}}(\overrightarrow{ind_0}));$

VI. CONCLUSION

The proposed system is restrained to the content of the knowledge base. It is important to notice that the system should work as kind of antiviral software. To make it usable one should be monitoring continuously the Internet and the steganographic software.

In the age of information and international terrorism we can only imagine how important is not to ignore the space of possibilities, which gives steganography both for good and evil goals. I emphasis that covert channels really exist even if they are not visible and even if it is impossible to prove they exist.

ACKNOWLEDGMENT

The author thanks Prof. A. Kraśniewski, the head of the PhD studies in the Institute of Telecommunication in the Faculty of Electronics and Information Technology of The Warsaw University of Technology. The author also thanks the promoter Prof. A. Dąbrowski for help in preparing theoretical basics for the subjects discussed in this paper.

REFERENCES

[1] Robert N. McDonough and Anthony D. Whalen, Detection of Signal in Noise, Academic Press, 1995.

[2] J. Fridrich, J., Goljan, "Practical Steganalysis of Digital Images - State of the Art", Proceedings of SPIE Photonics West 2002: Electronic Imaging, Security and Watermarking of Multimedia Contents IV, 4675:1-13, 2002.

[3] A. Westfield, A. Pfitzman, "Attacks on Steganographic Systems", Information Hiding, LNCS 1768, pp. 61-76, Springer-Verlag Berlin Heidelberg, 1999.

[4] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems", 3rd International Workshop on Information Hiding, IH'99 Dresden Germany, October Proceedings, Computer Science 1768. pp. 61- 76, 1999.

[5] A. Westfeld, A., Pfitzmann, A., "Attacks on Steganographic Systems - Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S-Tools - and Some Lessons Learned", Lecture Notes in Computer Science, 1768:61-75, 2000.

[6] N. Provos, P. Honeyman , "Detecting Steganography Content on the Internet", CITI technical report, vol. 1, pp. 01-11, August 2001.

[7] R.J. Anderson and F.A.P. Petitcolas, "On the limits of steganography", IEEE Journal of Selected Areas in Communications, Special Issue on Copyright and Privacy Protection, May 1998.

[8] H. Farid, "Detecting Steganographic Message in Digital Images", Report TR2001-412, Dartmouth College, Hanover, NH, 2001.

[9] H. Farid, "Detecting Steganographic Messages in Digital Images", Technical Report, Hanover, NH: Dartmouth College, 2001.

[10] N. F. Johnson and S. Jajodia, "Steganalysis of Images Created Using Current Steganography Software", Lecture Notes in Computer Science, vol.1525, Springer-Verlag, Berlin, 1998, pp. 273 -289.

[11] Magdalena Pejas, "Detecting Steganographic Transmissions", Software Developer's Journal 9/2005.

[12] Magdalena Pejas, "Aggregated Hypothesis Testing for Steganalysis", Transactions on Enformatika, Systems Sciences and Engineering 8 October 2005, 2005 International Academy of Sciences, ISBN 975-98458-7-3.