International Journal of Engineering and Management Research e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-2 (April 2023)
https://ijemr.vandanapublications.com https://doi.org/10.31033/ijemr.13.2.30

OSN-Tagging Scheme and Its Steganalysis Realizing Using Machine Learning Model

Lekshmi R.Nair
Computer Science and Engineering, College of Engineering Cherthala, INDIA

Corresponding Author: lekshmi.r.nair@cectl.ac.in

Received: 15-03-2023 **Revised:** 27-03-2023 **Accepted:** 29-04-2023

ABSTRACT

Steganography is a type of art and steganalysis is that art fining.In this work we propose a machine learning model for steganaysis. An SVM(Support Vector Machine) – Classification model. Testing the model with the help of OSN(Online Social Network)-Tagging scheme. Facebook was selected from all amoung the OSN for OSN-Tagging. Machine classify the steg-algorithm's accuracy in percentage.

***Keywords*—** OSN-Tagging, Steganalysis, SVM, AES, Bit Pattern, Useability, Peek Threshold Value

I. INTRODUCTION

Online social networks are dedicated websites that enable users to communicate with each other by posting information, comments, messages, images, etc. [11]. The popularity of OSNs such as Facebook, Twitter, Google+,etc.is continuously growing, with Facebook the most popular OSN based on the number of active users (active users are users who have logged in to Facebook in the last 30 days) [12]. In the third quarter of 2012, the number of active Facebook users surpassed 1 billion, while as of the third quarter of 2016 the number of active Facebook users have grown to 1.79 billion [13].

Steganography is the practice or art of hiding information in digital object [14], with image being the most popular choice of cover object [15]. Steganography's main objectives are undetectavility (resistance against both visual as well as statistical analysis) [16] . Although all three these objectives are desirable, most applications can only focus on one or two of these objectives and a trade-off is usually necessary. The main focus of the OSN-Tagging scheme that is on robustness, specifically against the types of image modifications that are performed by OSNs, and resist from statistical analysis (like entropy) that are analyzed by this machine learning methodology.

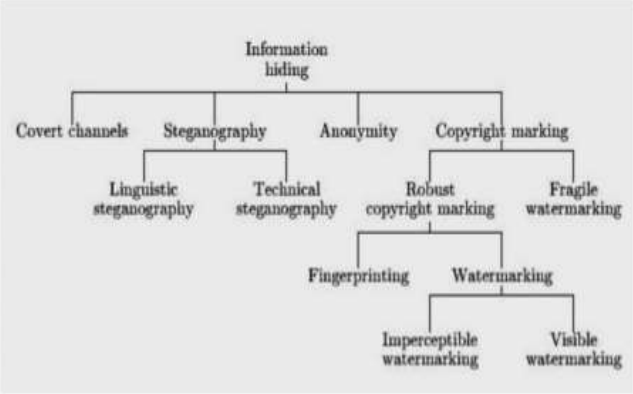


Figure 1: Classification of information hiding

Fig 1 represents the classification of information hiding. It can be mainly four type which are convert channel, steganography, anonymity and copyright marking. Watermarking is under copyright marking.

II. LITRATURE REVIEW

A. Advanced Encryption Standard(AES)

It is a symmetric block cipher. A number of AES parameters depend on the key length. The AES standard states that the algorithm can only accept a block size of 128 bits and a choice of three keys-128,192,256 bits. At present the most common key size likely to be used is the 128-bit key then the number of round is 10.

B. Support Vector Machine(SVM)

Support Vector Machine (SVM) is one of the most popular Machine Learning Classifier. It falls under the category of Supervised learning algorithms and uses the concept of Margin to classify between classes. It gives better accuracy than K-Nearest Niebour(KNN), Decision Trees and Naive Bayes Classifier and hence is quite useful.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-2 (April 2023)
https://doi.org/10.31033/ijemr.13.2.30

Figure 2: Classification of steganography

III. RELATED WORK

A. Mp3stego Steganalysis

Mp3stego is an open source data hiding algorithm that is built on top of 8 Hz encoder [6] . It means that files embedded with mp3stego will shares some characteristics with 8 Hz encoder. The algorithm embeds bits of message as the parity of *part2_3_length* field of SI. Embedding algorithm works directly on uncompressed samples of cover and embeds the message during the compression process. To that end, the algorithm adds a second criterion to inner loop of mp3 compression algorithm. Such that, not only the existing bit budget should be enough for encoding the granule, but also parity of its *part2_3_length* should match with bit of the message. Therefore, if parity of *part2_3_length* does not agree with the message, the inner loop is executed again and value of *global_gain* is changed [10] .

B. Caronni's Tagging

The OSN-Tagging scheme uses Tagging, one of the earliest watermarking schemes developed by Caronni [8] to protect and authenticate digital images. The Tagging scheme consists of adding small, geometric patterns to an image at brightness levels that are imperceptible to the human eye [9].The original goal of the Tagging watermarking scheme was to detect the source of illegal copies of data by inserting a different tag stream in each distributed image and matching the tags to a known list should the image be distributed illegally [8].

During the Tagging process, a series of NxN rectangles, called tags, are first identified [8]. The brightness level of each of the identified tags is adjusted, either increased or decreased, in such a way that the adjustment does not introduce visual artefacts to the image [8].

Table 1: Example of literature survey summery

Topic	Author Name	year	Description
Secure Data Transfer: Based on Steganography and Visual Cryptography	Rini K, D.Rajapriya	2017	OTP encryption+LSB+key image+stego image VC
An Efficient Secure Data Transmission Based on Visual Cryptography	Rubeena Jabi et.al	2017	TEA algorithm+Steganography +VC
Sharing a Secret Image with Encapsulated Shares in VC	Shankar K, Eswaran P	2015	VC+AES encapsulation

IV. PROPOSED WORK

A. Dataset

Use 50 plane image and its 50 stego-image.Choose 50 different types of images of different size, like nature pictures, animal, bird, flowers pictures. https://unsplash.com/s/photos/natural

B. Metric Used

1) **Homogeneity*:*** A material or image that is **homogeneous** is uniform in composition or character (i.e. color, shape, size, weight, height, distribution, texture etc.)
2) **Disimilarity-Coefficiant:** It is a statistic used to gauge the similarity of two samples.
3) **Entropy:** The **entropy** quantifies the amount **of** information needed to describe the outcome **of** a random variable Y. The entropy S is the natural logarithm of the number of microstates, multiplied by the Boltzmann constant k_B.

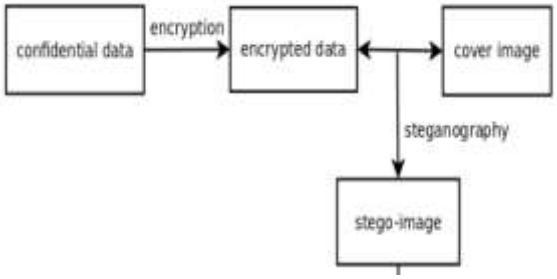
$$S = kB\ ln\Omega$$

C. Libraries Used

- Numpy
- Pandas
- Matplot lib
- Cv2
- Pyplot
- Glob
- Seaborn

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-2 (April 2023)
https://doi.org/10.31033/ijemr.13.2.30

Figure 3: Process involved in steganography



From Fig. 3 discribes the process involved in steganography. Confidential data with the help of encryption become encrypted data. The encrypted data hides in a cover image, the process is known as steganography. Which gives stego-image as output.

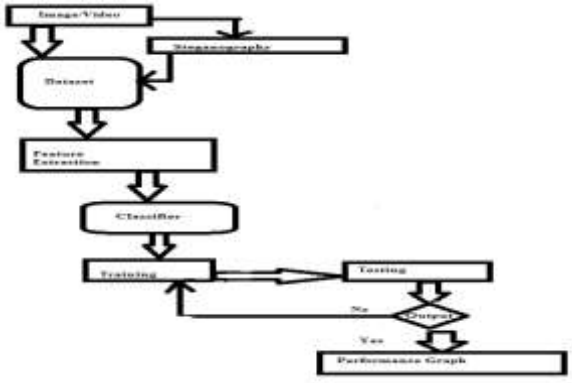


Figure 4: Architecture Diagram Design

This work is implemented as two parts. First part is the implementation of OSN-Tagging, Second is implementation of machine learning model. Support vector machine classification model is used here. Plane image is taken as x and its steganography applied image is taken as x' so both x and x' are going to be the input of ML-model. Then feature extraction taking place. Mainly finding glcm,homogeneity, enegy, entropy, mean these statistical features. Represented it in a 82 raws and 38 columns matrix known as feature matrix. This is the newly obtained dataset. Next is labelling using ones and zeros. Get 82 ones and zeros. Finaly calculation of confusion matrix(2raws and 2column). With the help of confusion matrix(below Table 2) calculating model accuracy in terms of percentage which is the final result. Model accuracy and stregthness of stego-algorithm are opposite in nature. Which means model accuracy is higher obviously strengthness of algorithm is lower. Aim is that find high strength steganographic algorithm by steganalysis and less model accuracy. Cofusion matrix contains

- True positive
- True negative
- False positive
- False negative

Table 2: Example of confusion matrix

Confusion Matrix1		Confusion Matrix2	
8	2	9	6
1	11	9	4
Confusion Matrix3		Confusion Matrix4	
7	6	5	8
11	4	3	10

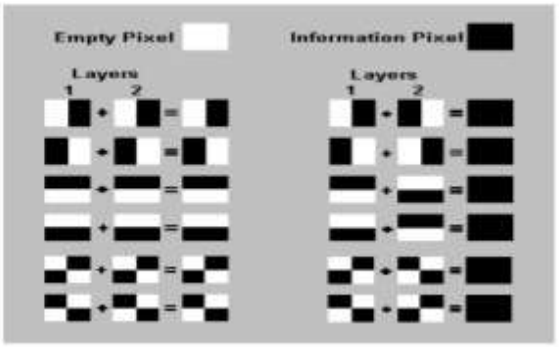


Figure 5: Secret data pixel distribution

Fig. 5 shows an example of pixel distribution of information. The various methods developed was to increase the number of shares. Multiple secret hiding scheme. pixel expansion where the pixel is expanded i.e the above discussed method. In order to prevent the intruder from noticing the shares. The concept of innocent looking shares were brought where the images are split in to innocent looking shares which is hardly noticeable.

- **Algorithm for Embedding**

```
tag-width□tag-height□10
bit stream=convert-to-bits(secret)
wr□5
wc□5
rows□image.height
cols□image.width
grayimage□convert-to-grayscale(image)
vmat□find-variance(image,grayimage)
tag-locs□find-tag-locations(vmat,wr,wc)
tag-mark□is-taggable(tag-locs)
tagh□rows/tag-height
tagw□cols/tag-width
locs=( )
for  i=0 upto tagh
for  j=0 upto tagw
if  tag-mark[i,j]=True
```

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-2 (April 2023)
https://doi.org/10.31033/ijemr.13.2.30

```
locs.append(i,j)
end if
end for
end for
for  each bit in bitstream :
(x,y )= remove(locs)
If bi t= =1
Img = increase-intencity(img,x,y)
Else
Img = decrese-intencity(img,x,y)
End  if
End for
Return  img
```

- **Algorithm for Recovery**

```
tag-height□tag-width□10
rows□image.height
cols□image.width
dif-matrix=orginal-image
tagh=rowstag-height
tagw=cols/tag-width
for  x=0 to tagh
x=x*tag-height
for  y=0 to tag-width
y=y*10
cell=extract-cell(x,y,tagheight,tagwidth)
mean=calculate-mean(cell)
if  mean< -1;
bits append(1)
else
bits append(0)
end if
end for
end for
secret-convert-to-string(bits)
return secret
```

Figure 6: Example of input image

Figure 7: Example of an image with suitable tag locations

In the above Fig 7 shows suitable tag locations. There are around 20 suitable tag locations found.

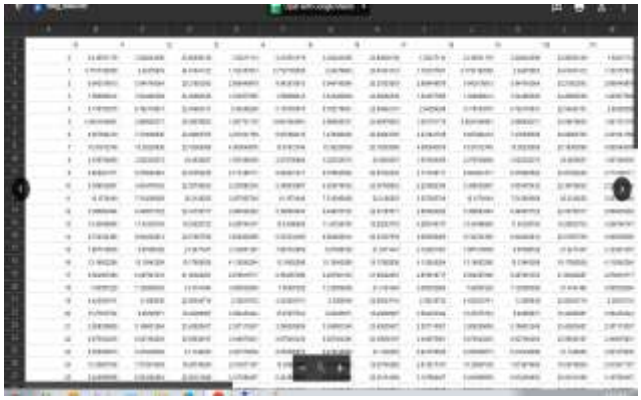


Figure 8: Example of extracted dataset representation

Above Fig 8 is an example of extraction dataset representation. After feature extraction new dataset representation is like this.

V. RESULTS AND OBSERVATIONS

Above table2 gives the variable size text and same picture size encryption and decryption time. Table3 gives the encryption and decryption time of fixed text size and variable picture size. Analysis gives in both cases not harmly effected the encryption and decryption time.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-2 (April 2023)
https://doi.org/10.31033/ijemr.13.2.30

Size Of Text	Picture Size	Encryption Time	Decryption Time
18 bytes	28.5 KB	20.4 s	12.2 s
420 bytes	28.5 KB	20.2 s	11.7s
1366 bytes	28.5 KB	20.1	12 s

Table: Encryption and Decryption Time when Variable Input Size Fixed Cover Size

Size Of Text	Picture Size	Encryption Time	Decryption Time
18 bytes	28.5 KB	20.4 s	12.2 s
18 bytes	38.2 KB	20.4 s	11.9s
18 bytes	45.6 KB	19. 8 s	11.8 s

Table: Encryption and Decryption Time when Fixed Input Size Variable Cover Size

Figure 9: Result Analysis 1

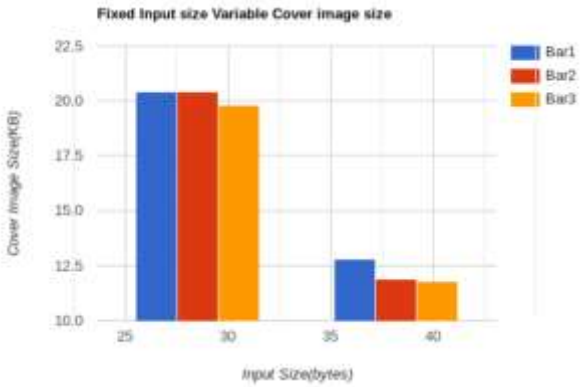


Figure 10: Result Analysis 2

From Fig 9 & 10 we can say that both graph are same. This is just the graphical representation of table 3&4.

VI. CONCLUSION AND FUTURE SCOPE

In OSN uploaded images are struggle with image processing attacks and undetectability of information this paper propose a solution with maximum payload capacity, And also proven by steganalysis of OSN-Tagging scheme while designing a machine learning model.

Future work can be done to strengthening the statistical undetectability of OSN-Tagging.

ACKNOWLEDGMENT

The author would like to thank god and conformed that the work is genuine.

REFERENCES

[1] Tayana Morkel,"The OSN-Tagging Scheme:Recoverable Steganography for Online Social Networks",2017

[2] Lekshmi.R. Nair, Aggie Varghese and Suresh Kumar .N"A Smart Steganography and Recovery for Online Scial Networks",2020

[3] Sahar A. El_Rahman," A comparative analysis of image steganography based on DCT algorithm and steganography tool to hide nuclear reactors confidential information",2016

[4] Hemalatha S, U. Dinesh Acharya, and Renuka A"Wavelet transform based steganography technique to hide audio signals in image",2015

[5] Amine Benhfid, El Bachir Ameur and Youssef Taouil,"Reversible steganographic method based on interpolation by bivariate linear box-spline on the three directional mesh",2018

[6] Hamzeh Ghasemzadeh," Calibrated steganalysis of mp3stego in multi-encoder scenario",2019

[7] Rubeena Jabi,"An Efficient Secure Data Transmission Based on Visual Cryptography",2016

[8] G. Caronni, "Assuring ownership right for digital images" in Proceedings of the Conference on Reliable IT Systems, 1995, 251-263.

[9] I.J. Cox, J. Kilian, F.T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia" in IEEE Transactions on Image Processing, 1997, 6(12):1673-1687

[10] F. Petitcolas, mp3stego, Http//Www. Cl. Cam. Ac. Uk/ ~Fapp2/Steganography/Mp3stego/Index. Html. (1998).

[11] Oxford English Dictionary, Oxford University Press, 1989

[12] Leading social networks worldwide as of September 2016, ranked by number of active users (in millions) http://www.statista.com/ statistics/272014/global-social-networks-ranked-by-number-ofusers/,last accessed on 12 January 2016.

[13] Number of monthly active Facebook users worldwide as of 3rd quarter 2016 (in millions), http://www.statista.com/statistics/ 264810/number-of-monthly-active-facebook-users-worldwide/, last accessed on 12 January 2016.

[14] J. Fridrich, Steganography in digital media: Principles, algorithms and applications, Cambridge University press, 2010

[15] T. Morkel, J.H.P. Eloff and M.S. Olivier, "An overview of image steganography" in Proceedings of the Information Security South Africa (ISSA) Conference, 2005.

[16] A. Cheddad, J. Condell, H. Curran and P. McKevitt, "Digital image steganography: Survey and analysis of current methods" in Signal Processing, 2010, 90(30):727-752.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-2 (April 2023)
https://doi.org/10.31033/ijemr.13.2.30

[17] Wikkipedia-The encyclopedia,2018
[18] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
[19] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
[20] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
[21] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[22] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
[23] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/
[24] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
[25] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
[26] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
[27] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
[28] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.
[29] *Homer Benny Bandela,M.Ganesh,Babu et .al"Crypto-stego Technique for Secure Data Transmission."2019.*
[30] *Tribastuti Yuniati and Rinaldi Munir"Secure E-payment Method Based on Visual Cryptography "2018.*
[31] *Atanu Sarkar and Sunil Karforma"I mage Steganography Using Password Based Encryption technique to Secure E-pay"2018.*
[32] *Shankar k and Eswaran p,"Sharing A Secret Image with Encapsulated Shares in VC"2015.*
[33] Bikash Dutta. (2015). Rural development through self help groups: An overview. *Indian Journal of Applied Research, 5*(4), 70-78.
[34] Wilkie, W. L. (1994). *Consumer Behaviour*. (3rd Edition). New York: John Wiley and Sons.
[35] Saunders C. S. (2014). Point estimate method addressing correlated wind power for probabilistic optimal power flow. *IEEE Transactions on Power Systems, 29*(3), 1045-1054.