Email Security using Hybrid Approach and Classification

Anjaly Ezhava
Computer Science and Engineering Dept.
R.N.G. Patel Institute of Technology
Bardoli, India
anjalibiju90@gmail.com

Palak Varu
Computer Science and Engineering Dept.
R.N.G. Patel Institute of Technology
Bardoli, India
palak.varu2000@gmail.com

Kamya Rathod
Computer Science and Engineering Dept.
R.N.G. Patel Institute of Technology
Bardoli, India
kamya1622@gmail.com

Bhavini Bhatt
Computer Science and Engineering Dept.
R.N.G. Patel Institute of Technology
Bardoli, India
brb.fetr@gmail.com

Hina Chavda
Computer Science and Engineering Dept.
R.N.G. Patel Institute of Technology
Bardoli, India
heenachavda106@gmail.com

***Abstract*–Today Email is considered to be the most important medium of communication for both professional and non-professional work. Also the attachment in email contains important documents, messages, files etc .The need for more secure and advanced dissemination is ever increasing since advancement in digital transmission has also posed equally challenging threats. Hence security becomes a major aspect in email. The technique Cryptography converts the data into cipher text so that original data will not be in readable form to others and to hide the conversation, steganography is the process of hiding a secret message within an ordinary message and extracting it at its destination Also, it becomes easy for users if the mails are organized in professional and personal categories. Hence in our proposed system, we will provide security through steganography andcryptographyandprovideclassificationforpersonalandpro-fessionalmails.**

Keywords–steganography, cryptography, email, security, classification

1 INTRODUCTION

With the expansion in digital communication technologies, email has become the modern means of communication and plays a significant role in our life. With this, the necessity of securing it withmorepowerfulsecuritybecomesmandatory.Mostofthenamed organizations use TLS SSL to secure their system. And to make it more secure, the idea to implement technologies viz.Cryptography and Steganography comes to mind. Cryptography transforms data into seemingly meaningless bits, called ciphertext, by using a sophisticated and robust algorithm. This will help the intended recipient to recover the original message by means of a crypto-graphic key. For those who do not have the key, the encrypted message will appear as a stream of meaningless codes. But still the communication or data transfer is visible to others. To overcome this, Steganography is used to camouflage the presence of the hidden data in such a manner that no one other than the sender and the proposed recipient even recognizes that here is a hidden message. The information is hidden inside a message so that the third person can just see the image and the data will not be visible. The principle involves two prime components-A secret or hidden message and the cover, which is a medium, where the secret is hidden. The hidden message can be a piece of text, a random image, speech or even a full video but for now we are only considering text. A stego or container is generated which contains the hidden content inside the cover. The hidden content can be extracted from this stego. In this paper, the type of steganography used is image steganography, where the cover is an image. In image steganography, we are going to consider the secret information as a piece of text. As email is one of the most used among users, filtering email based on a particular property is also essential. The process of filtering email is known as Classification in terms of machine learning. There are several algorithms used to classify/filter emails. These algorithms will be tested on the basis of the accuracy it provides. So, we will be considering data sets to measure the accuracy of a particular algorithm. This paper is organized in five sections. Section I gives an introduction of the modules in this paper. Section II discuss the related work of cryptography, steganography and mail classification. Section III displays the proposed methodology of cryptography, steganography and Classification. Section IV briefs about Results and Discussion of the proposed methodology. Section V gives the conclusion of this paper.

2 RELATEDWORK

In [1], authors say that Random Forests for Spam base dataset provides better accuracy when compared to other Machine Learning techniques like Decision Tree, Na¨ıve Bayes, and Support Vector Machine. In [2], authors concluded that the Multinomial Na¨ıveBayesgivesthebestoutcomebuthaslimitationsduetoclasscondi-tional independence which makes the machine to misclassify some tuples. Ensemble methods including “Random Forest” on the other hand have proven to be useful as they use multiple classifiers for class prediction. In [3] , authors concluded that the size of the data to be hidden depends on the size image used and the speed of the medium of the carrier is also one of the most important factors that we must consider to facilitate transfer which will reduce the amount of data transferred. The lossy compression has the largest proportion compared to the lossless compression. This is because very small file size and a lot of tools, plugins and software support the largest proportion. In addition, the reluctance of some researchers because of quality degrades with the highest propor-

tion of compression that makes it hard to return to its original size After the compressing process. Meanwhile, the lossless compression is low among some researchers who believe that it is actually not a loss of quality last rather a slight decrease in image file sizes. In[4],authors have conducted an experimental study on the impact of selected steganography methods on JPEG file size. Overall, our experiments show that different steganography methods have different impacts on the file size. F5, nsF5, and PQ reduce the file size while OutGuess 0.1, OutGuess 0.2, complementary embedding increase the file size. JUNIWARD and SIUNIWARD preserve the file size of the cover image. In [5], As Steganography only hides the data into another medium (for e.g. digital image), once noticed the hidden data could be retrieved with known algorithms. Thus Steganography must be used in conjunction with Cryptography to combine data obfuscation and data hiding properties to make Message Communication between Sender and Receiver even more secure. However, Steganography alone gives the clear advantage over cryptography such that messages do not attract attention to themselves, to messengers, or to recipients. In [6], AES and 3DES cryptography method successfully implemented on the email text message in order to have a better security level. The recovery process after the sending process also was successfully done. Evaluation's results show that AES is better in terms of compile time, while 3DES is better in terms of increasing message's size after the encryption process. But because the change in the addition of bytes is not significant, it does not have an impact on the duration of sending messages through ESP. Based on some of the results of these tests, email users are recommended to use AES encryption. In [7], on the transmitter side, the picture in addition to the message is enciphered simultaneously. On the sender side, first the data is enciphered using the AES algorithm, now this enciphered confidential data is embedded into the picture using modified BPCS, as it has abundant data hiding capacity. Then this picture is enciphered using the AES algorithm. Digital watermarking technique is used on the enciphered stego picture. This watermarked enciphered stego picture contains confidential data which is then passed over the internet to the retriever. On the retriever side, the watermarked picture is dewatermarked. It analyzes in case there are any modifications in the watermarked picture or not. In [8], it is mentioned that the image steganography methods are of two types depending on domain type: spatial domain based techniques and frequency domain based techniques. In spatial domain based technique, the message is embedded in the intensity of pixels of the images straightly while in frequency domain based technique, images are converted into the frequency domain and then the messages are embedded in the transform coefficients. Among many spatial domain based techniques, LSB (Least Significant Bit) method is the widely applied method. In [9], The stego picture comprisesthedataprogrammedintheLSBoftheimage.TheLSBcentredsteganography is shared among shifting algorithms to improve the safety level of the image. Shifting algorithm modifies the pixel positions of the stego picture thus the secreted information might notberecoveredeasily.Thestegoimagecomprisesthesecretinformationthatcanbeeasilysentoverawirelessnetwork.Theintruder, even if he gets access to the image file, would not know that the data is hidden in it. In the past one year, with the rise in research in deep learning, Steganography and Steganalysis have emerged as a prime application. [10] has further worked on it, with introducing a

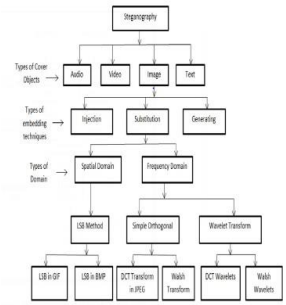

Fig.1 Classification of Image Steganography

Competitive coevolution learning approach wherein Generative Adversarial Networks(GANs) are used for performing steganography along with steganalysis to improve the overall results. RGB color images are used as cover and secret images. [11] proposes an embedding technique based on cyclic chaos, which effectively embeds secret messages into color images. The cyclic chaos is divided into two parts: the chaotic function of generating seeds and the pseudorandom number generator (PRNG). A pixel is found through the chaotic cycle system, and a three-bit secret message is embedded into the R, G, and B channels of the pixel through a specific embedding algorithm. Compared with other color image embedding methods, the proposed algorithm has higher visual quality and improved security. In [12], the LSB method is proposed for image steganography with .bmp image format because it uses lossless compression. For cryptography, AES technique is proposed which uses 256-bit symmetric key for encryption. In [13], the IntelligentwaterDropletalgorithmiscombinedwithNaive-Bayesalgorithm
.Also it is considered better than ACO and GA algorithms. In [14], the author proposed a method for hiding the data in color cover image. For that starting position in the image and the message length (row, column, length) is defined. This position can be used as a first secret private key (key1) and insert character of message by reserving 1 byte of image for character. Reshape 3D to 2D matrix and divide in equal blocks(4*4) and use it as a secret key(key2). Apply XOR to encrypt 2D matrix(with key2),reshape 2D to 3D matrix for encrypted color image. In [15], to hide data in cover image LSB technique is used. It is also mentioned that the payload capacity of LSB technique is very high in which first the secret data convert into the binary bits and each binary bits are replaced by the LSB of cover image. Other methods like DCT and DWT are also com-pared and concluded that LSB is convenient amongst all. And for cryptography RSA method is proposed.

3 PROPOSED METHODOLOGY

3.1 Proposed Email with hybrid Security Architecture

The architecture of email with hybrid security shown in figure 1. Is an email system where two users communicate through emails and are provided with security options viz. Cryptography and Steganography while composing mails. Also the received mails get classified into two sections Personal and Professional.

First phase in the system is the user needs to register into the system if not or if already registered then can access the system by logging into it. Then in the second phase the user can compose mail where he/she is provided with security options i.e. Cryptography - which will encrypt the message that he/she wants to send and Steganography - which will hide the message into an image so that the message will not be visible and send mail to the respective user. And if the user wants to view mails then can view in the classified sections i.e. Personal - where personal mails from friends or any personal mails gets received and Professional - where mails from work place , events or received for any professional work get received.

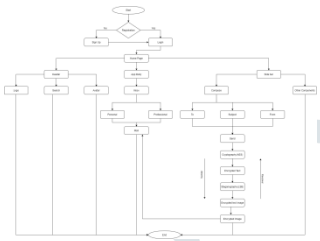

Fig.3.1 System Flow of Proposed System

3.2 Cryptography module to encrypt data

In cryptography, the message is encrypted so that it becomes unreadable to others. As a result of this data encryption, the hacker is unable to access the email content. Therefore, your outgoing messages remain safe until they reach the recipient. Below figure shows the working of cryptography

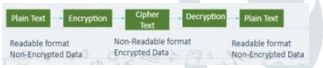

Fig.3.2 Working of Cryptography

3.2.1 AES Algorithm.

There are many algorithms for cryptography but for our proposed system we will be considering AES (Advanced Encryption Standard). AES is an iterative, based on 'substitution permutation network'. AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plain text block as 16 bytes. The number of rounds in AES is variable and depends on the length of the key.AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES

3.3 Steganography module to hide the data

While cryptography provides privacy, Steganography is intended to provide secrecy. Hiding information in other information when communication is taking place is usually referred as steganography. And to hide data digital images are the most popular among different file formats available on internet because of their frequency.

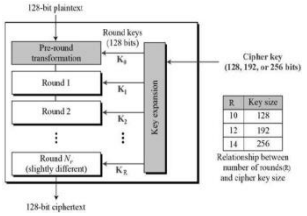

Fig.3.2.1 AES Structure

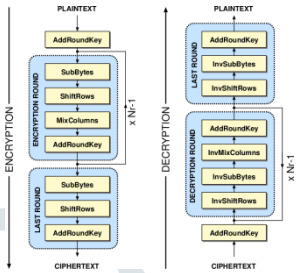

Fig.3.2.2 AES Rounds

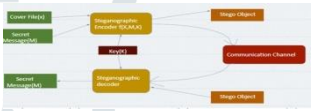

Fig.3.3 Working of Steganography

3.3.1 Least Significant Bit (LSB).

In this method the least bit of an image is altered and the data to be hidden is stored in those bits. Since it is called least significant bits so by changing those bits in an image the image view is not tampered. Inside every image there is RGB color combination, we take one pixel of the image of 3 bytes for Red and Green and Blue of 8 bits each. Every byte the last bit is taken out and data is put in these. The data to be put is first converted into binary. LSB uses .bmp images because they use lossless compression. LSB method explained by an example below:

Grid of 3 pixels of a 24 bit image can be written as

(001011010001110011011100)

(101001101100010000001100)

(110100101010110101100011)

When 200 whose binary representation is 11001000 is embedded in the least significant bits of the part of the image, the resulting grid is as follows:

(001011010001110111011100)

(101001101100010100001100)

(110100101010110001100011)

The alteration done by editing the least significant bits of the image to put in the message is not visible to the naked eye.

Below figure shows the practical working of steganography

Fig.3.3.3 Steganography Process

3.4 Classification

In basic words, classification is the action of categorizing into separate groups. With an increasing number of users, identifying emails into different categories becomes our necessity

3.4.1 Naïve Bayes.

Naïve Bayes is one of the utmost well known algorithms applied in these procedures. Na¨ıve Bayes classifier was used in 1998 for spam recognition. The Na¨ıve Bayes classifier algorithm is an algorithm which is used for supervised learning. The Bayesian classifier works on the dependent events and works on the probability of the event which is going to occur in the future that can be detected from the same event which occurred previously. Na¨ıve Bayes was made on the Bayes theorem which assumes that features are autonomous of each other. Na¨ıve Bayes classifier technique can be used for classifying emails as word probability plays the main role here. Naïve Bayes classifier algorithm has become the best technique for email filtering. For this the model is trained using the Na¨ıve Bayes filter very well to work effectively. The Naïve Bayes always calculates the probability of each class and the class having the maximum probability is then chosen as an output. Na¨ıve Bayesal ways provide an accurate result. It is used in many fields like filtering.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$$P(B) = \sum_{y} P(B|A)P(A) \quad (2)$$

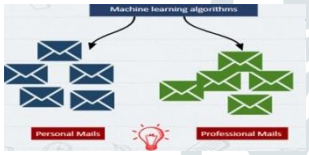

Fig.3.4.4 Classification of Emails

4 RESULTSANDDISCUSSION

4.1 Operation of email hybrid system

System consists of different modules like registration/login, email compose, choosing cryptography and steganography and mail classification. The result for the same are discussed below:

4.1.1 Registration and login.

If the user is new then first he/she needs to register into the system. Once registered, a user can access their account by logging into it. It requires user email address and password for the account accessibility.

Below figure shows the registration and login for our system.

4.1.2 Compose the email.

To compose an email, a user needs to give the email address of the user to whom he/she wants to send the email. After that the subject and message field is needed to be filled. There are options available to users if he/she wants to use either steganography, cryptography or both at the same time. After this user can send the email and at the other end receiver will receive the email.

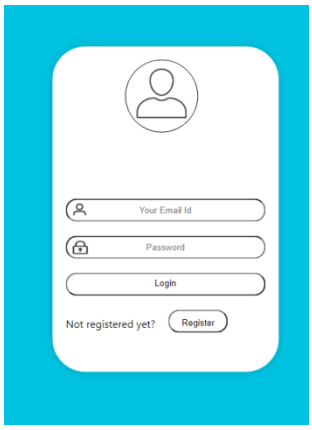

Fig.4.1.1 Registration and Login

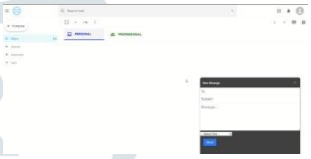

Fig.4.1.2.1 Compose Email

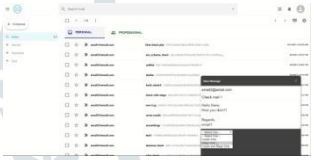

Fig. 4.1.2.2 Compose with 3 different security options

4.1.3 Classification of email.

When an email is received, it gets classified into two sections Personal and Professional with the implementation of Naïve Bayes Algorithm. A user can view the emails according to the particular section which he/she is looking for.

Below is the screenshots of two tabs for Personal and Professional emails of our system.

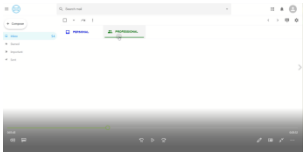

Fig.4.1.3.1 Personal email classification

4.1.4 Received emails with Steganography and Cryptography applied.

As seen in figure, the received email is either hidden inside image(whensteganographyisselectedatsenderside)orencryptedmessage(whencryptographyisselectedatsenderside)orboth

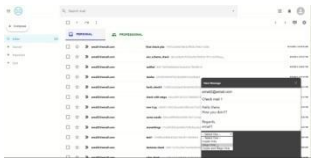

Fig.4.1.3.2 Professional email classification

encrypted and hidden inside image(when steganography and cryptography both selected at sender side). The Decrypt Message Button will show the decrypted messages in popup window.

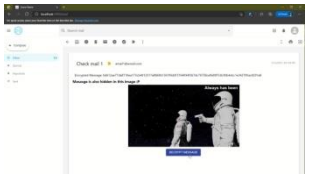

Fig.4.1.4.1 Received email with Steganography

Fig.4.1.4.2 Decrypted message from stego image

Fig.4.1.4.3 Received email with Cryptography

5 ACKNOWLEDGMENT

We wish to express sense of true gratitude towards Prof. B. R. Bhatt who at every discrete step in preparation of this paper contributed her valuable guidance and help to solve every problem. With all respect and gratitude, we owe our success to the writers of the reference papers that are referred by us in this paper activity.

6 CONCLUSION

The implementation produces the secure email system with Cryptography technique called AES (Advanced Encryption Standard) which is faster and secure than other encryption techniques. For image Steganography, LSB (Least Significant Bit) is considered which is most convenient for image steganography. For email classification for professional and personal email, we have considered Naïve Bayes algorithm which is considered as the best technique for email classification.

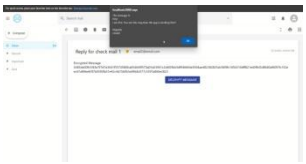

Fig.4.2.1.2 Decrypted message from encrypted data

7 FUTURE WORK

Upon successful implementation of the proposed project, we would integrate speech-to-text feature in the email, along with steganography supporting various formats of file . This will enable more ease and user satisfaction.

8 REFERENCES

[1] V. Sri Vinitha , D.Karthika Renuka “Performance Analysis ofE-Mail Spam Classification using different Machine Learning Techniques” 2019 International Conference on Advances in Computing and Communication Engineering(ICACCE).

[2] Nikhil Kumar , Sanket Sonowal, Nishant “Email Spam Detection Using Machine Learning Algorithms” 2020 Second International Conference on Inventive Research in Computing Applications(ICIRCA).

[3] Roshidi Din, Osman Ghazali, Alaa Jabbar Qasim “Analytical Review on Graphical Formats Used in Image Steganographic Compression” Indonesian Journal of Electrical Engineering and ComputerScience2018.

[4] Mohammad Rezaei , Saeed Montazeri Moghadam “Impact of Steganography on JPEG File Size” 27th Iranian Conference on Electrical Engineering(ICEE2019).

[5] Ms. B. Veera Jyothi, Dr. S. M. Verma, Dr. C. Uma Shanker “Implementation and Analysis of Email Messages Encryption and Image Steganography Schemes for Image Authentication and Verification” International Journal of Computer Applications (0975–8887)Volume5–No.5,August2010.

[6] Rini Indrayani , Subekti Ningsih, Pramudita Ferdiansyah , Dhimas Adi Satria “Effectiveness comparison of the AES and 3DES cryptography methods on email text messages” 2019 International Conference on Information and CommunicationsTechnology (ICOIACT).

[7] Sakshi Audhi, Maruska Mascarenhas —”Secure Mechanismfor Communication Using Image Steganography” , International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT),2019.

[8] Chitra Biswas , Udayan Das Gupta, Md. Mokammel Haque - ”An Efficient Algorithm for Confidentiality, Integrity and Authentication Using Hybrid Cryptography and Steganography” ,2019 International Conference on Electrical, Computer and Communication Engineering (ECCE),7-9February,2019.

[9] Seema Chavan , Dr. Y B. Gurav - ”Lossless Tagged Visual Cryptography Schemeusing Bit Plane Slicing for Image Processing” , Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA2018).

[10] Sarthak Ahuja, C. Udaya Kumar, and Hemalatha - “Competitive Coevolution for Color Image Steganography” ,Proceedings of the

International Conference on Intelligent Computing and Control Systems (ICICCS2019).
[11] Jiayu Deng, Mingwei Tang, Yanting Wang, Zhen Wang-" LSB Color Image Embedding Steganography Based on Cyclic Chaos",2019 IEEE 5th International Conference on Computer and Communications.
[12] Uzair Nisar, Craig Stewart, " Implementation of Email System With Steganography" International Journal of Computer SciencesandEngineering,Vol.7,Issue.1,pp.168-173,2019.
[13] Singh, Maneet. " Classification of spam email using intelligent water drops algorithm with naive bayes classifier. "Progress in Advanced Computing and Intelligent Engineering. Springer, Singapore, 2019.133-138.
[14] Rasras, RashadJ.,Ziad A. AlQadi, and Mutaz Rasmi Abu Sara."A methodology based on steganography and cryptography to protect highly secure messages." Engineering, Technology and Applied Science Research 9.1 (2019):3681-3684.
[15] Kumar, Ravi, and Namrata Singh. "A Survey Based on Enhanced the Security of Image Using the Combined Techniques of Steganography and Cryptography." Available at SSRN 3563571(2020).