A More Private & Secure E-Mail System using Image Steganography (EPS) and Data Mining

Ruchi Sharma
M.Tech Scholar
Department of Computer Science
TIT&S, Bhiwani
+91-9034915114
rsharma.ruchi92@gmail.com

Nidhi Sharma
Assistant Professor
Department of Computer Science
TIT&S, Bhiwani
+91-9466317526
nidhisharma1725@gmail.com

ABSTRACT
Data mining is a practice of automatically exploring and analysis of large quantities of data in order to discover valid, potentially useful and understandable patterns in data [1]. The data provided may contain private and user sensitive data leads to increasing concern about privacy and how to preserve it? Basically privacy preserving is an important issue in the field of data mining which deals with hiding individual's sensitive identity against unsolicited disclosure. There is also need of privacy preserving methods of communication which we are using regularly such as E-mail. E-mail is one of the most popular mode of communication due to its low cost, better usage of mails and business potentials. In order to discover user need and knowledge in mailing various data mining techniques have been applied on e-mail data to find unknown pattern. As email data also contains sensible information. In today's world people have become well aware of the privacy threats on their personal and sensitive data which is kept on e-mail is viewed anywhere. In this paper, we present a brief model for privacy and security on e-mail mining using a new designed steganography approach EPS(E-mail Privacy Steganography).

Keywords
E-mail; Mail mining; Security; Privacy; Steganography; EPS;

1. INTRODUCTION
There are some areas that must be known before discussing our model of privacy and security for mail. These are as follows:

1.1 Data Mining
Data mining has made broad significant multidisciplinary field used in vast application domains and extracts knowledge by identifying structural relationship among the objects in large data bases [1]. Data mining plays very significant role as it use to search from a large amount of data in order to find pattern.

1.2 E-mail and its Routing
E-mail stands for Electronic mails. It is a commonly and widely used platform for communication. In E-mail messages are transferred and received electronically. The both the receiving as well as sending parties are not need to be online at a same time. E-mail works with the help of three main protocols namely SMTP, POP, and IMAP.

SMTP stands for Simple Mail Transfer Protocol and is used for sending mail. The mails are sending from a sender machine to mail server of the sender. The mail server of sender delivers it to the mail server of recipient. The mail server of recipient further delivers the mail to recipient mailbox.

POP/IMAP stands for Post Office Protocol/ Internet Mail access Protocol. They are responsible for delivering the mail message to the recipient client. These protocols will further deliver these mails to mail client using these protocols.

1.3 Mail Mining
Mail mining is the procedure of analyzing content of e-mail messages and drawing useful patterns as a result. The procedure of automatic classification of e-mail according to its content, its author etc is also performed using mail mining. As a result of growing e-mail misuse problem, automated methods for analyzing the content of e-mail message are a vital requirement and this is performed with the help of mail mining techniques [2].

The procedure of mail mining is performed using clustering and classification techniques of data mining.

1.4 Privacy and its Requirement in E-mail
Due to rapid growth of internet and information technology, the ever growing privacy concern has become a major hurdle for information sharing. The darker side of showing personal information is that web user loses control over who can access to their information. Data mining plays very significant role as it use to search to large amount of mail data in order to find private mail. The goal of this is to find that information which is private for a user or group of user. When this information is found then they are further being secured in order to provide privacy of sensitive data.

Privacy is the one of most important concern in mail because if it is violated it may have adverse effect user It may be financially or socially or any other type of adverse effect. Privacy persevering data mining algorithm have been introduced with aim of preventing the sensible information which are containing in our E-mail. E-mail contains very much amount of private data. In order to ensure its privacy there is a need to provide privacy preserving algorithm like PGP. Our proposed model privacy preserving for e-mail is better and more secure in some aspects then other models like PGP. This proposed Model use Steganography. In this we implement an algorithm for steganography called EPS (E-mail Privacy Steganography) which is used to hide private data.

AICTC '16, August 12 - 13, 2016, Bikaner, India

ACM 978-1-4503-4213-1/16/08...$15.00
DOI: http://dx.doi.org/10.1145/2979779.2979827

1.5 Steganography

Steganography is the method of hiding personal data from intruder. It provides secrecy of text/images to prevent the data from attackers. It will embed the message over digital object such as image. Hence intended intruder will unable to detect the presence of message. There may exist a large no. of method of steganography. Least-Significant-Bit (LSB) technique is simple to understand and easy to implement. Image steganography has three major aspects [3].

1.5.1 Capacity

The data that can be stored in cover image should be maximum

1.5.2 Imperceptibility

The visual quality of the image after data hiding should be good.

2.5.3 Robustness

The procedure of hiding data should be robust

2. EXISTING TECHNIQUE (PGP)

The currently used common model for privacy preservation is PGP (Pretty Good Privacy). A large amount of work is done on this algorithm for e mail privacy preserving in the last decade to make it more secure.

Phil Zimmermann created the first version of PGP encryption in 1991. The name, "Pretty Good Privacy" was inspired by the name of a grocery store, "Ralph's Pretty Good Grocery", featured in radio host Garrison Keillor's fictional town, Lake Wobegon [5].In June 5, 2001, PGP Marks 10th Anniversary"[5].

PGP is a combination of both public and private (Symmetric key) cryptography. PGP firstly use a session key to encrypt the message, which is basically a onetime symmetric key. This key is used by a symmetric key algorithm to encrypt the message. Then this one time session (symmetric) key is then encrypted by recipient public key which is generated by use of RSA algorithm. Then the encrypted data along with encrypted session key is transmitted to recipient.

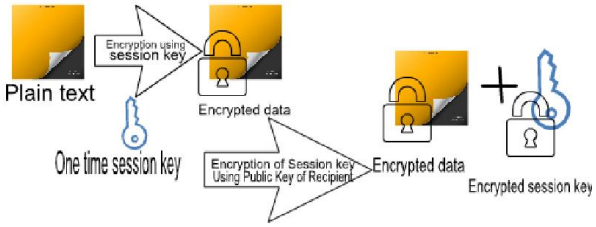

Figure 1. PGP Encryption.

The Decryption process is reverse of the encryption process i.e. the key is decrypted using recipient private key and then this session key is used to decrypt the message.

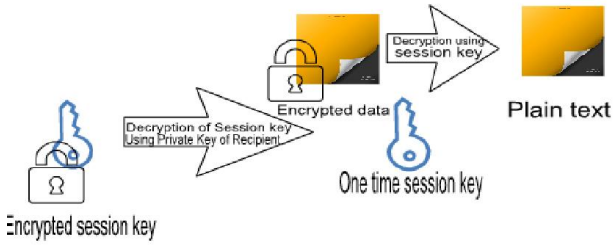

Figure 2. PGP Decryption.

Certain new versions of PGP are implemented to provide more security as early versions of PGP have been found to have theoretical vulnerabilities. These are PGP 2, PGP 3. The original version of PGP uses RSA approach where as the PGP 2 use symmetric key algorithm IDEA. While the current version of PGP uses some other technology for encryption.

The PGP is basically a cryptographic method to provide e-mail privacy. The PGP is a combination of both types of cryptography i.e. it uses the speed of symmetric key cryptography with a convenience of public key cryptography. Hence the PGP is faster than the normal public key encryption.

The PGP is a good method to achieve confidentiality and privacy as the message can't be decrypted so easily hence privacy of data is preserved. But the PGP includes following drawbacks which are need to be rectified.

2.1 Drawbacks of PGP as compared to EPS

But PGP algorithm has several disadvantages. These are as follows:-

2.1.1 Compatibility Issue

The PGP software of both sender and receiver must be of same or compatible version as the scheme of encryption is different for different version but EPS do not require such a problem.

2.1.2 Complexity

The complexity of PGP technique is a major issue as the encryption in PGP required complicated algorithms but this EPS algorithm is not complex to understand.

2.1.3 Key Handling

PGP requires two types of keys and handling of these keys is a major concern in PGP but in EPS no concept of key handling as key is presented in a USB device.

Our proposed model will solve all these problems and improve the security and privacy of mail and make it more private and secure.

3. PROPOSED MODEL

Our Proposed model uses the data mining and Steganography to provide the privacy and security of mail. The proposed secure mail system uses the separate models for sending and receiving. The sending and receiving of mails is possible between users of our mail systems only. There is a large pool of background images provided in our mail system.

This mail system has a secure mail client which uses password or some other biometric authentication depends upon the requirement of users. But initially it uses only password authentication. The mail client will connect to a separately established mail server. The system mines each and every mail to check whether it is confidential or not. The sending and receiving mail uses Steganography to hide the confidential mails hence the private data is kept secret. Hence we can say privacy is preserved.

3.1 Sending Mail and Encoding

The sending procedure will determine type of the mails and transfer it from sender's client to recipient server.

The Sending Procedure uses the proposed steganography algorithm EPS (E-mail Privacy Steganography) for sending private data.

The sender will firstly log in using the mail client installed in sender machine. The authentication method is used to log in. The

system will check the credentials of the user and determine whether the user is permitted to log in or not. If the credentials are right then the user can access corresponding mail account. The user then select compose mail option displayed in mail account. Then message is then typed in the compose mail dialog box. Then type of the mail is determined by data mining algorithms and if the mail is confidential then our proposed algorithm EPS (E-mail Privacy Steganography) and a key K is used to encode it in the background image of sender, The Key K will determine the pixel value from where the data hiding is starts. The key is kept per user wise. Rest of the mails are kept as it is i.e. the non confidential mails are kept unhidden. Then send background image (contains confidential mails) along with the unencrypted mails to the mail server.

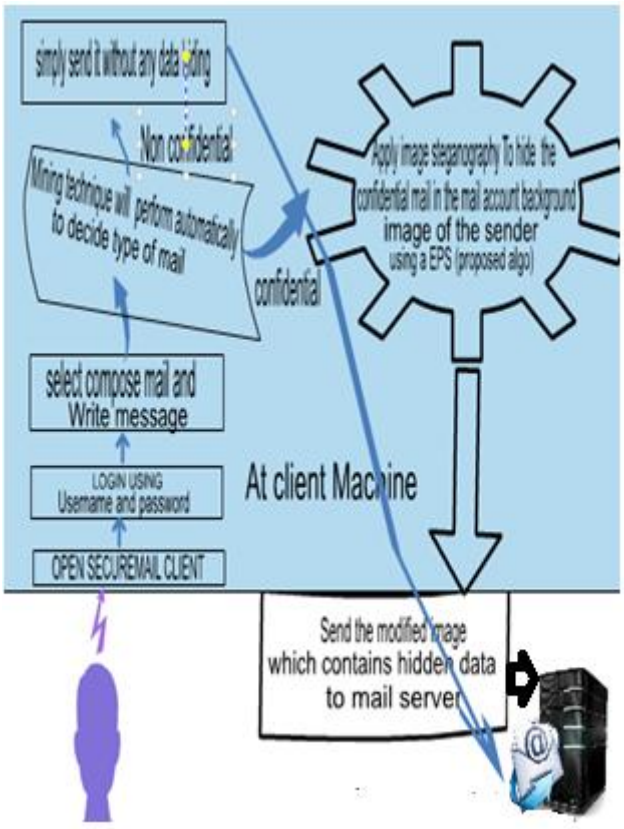

Figure 3. Sending mail procedure.

3.2 Receiving mails and Decoding

The mails are stored initially at the mail server only and these are downloaded in the recipient machine only when he is logged in.

The user will firstly log in using mail client installed in the user machine using authentication method (Biometric or Password). We are using the password authentication. The mail client will check the authentication details if these are correct then user is permitted to access the mail account. The mail client is connected with the user mailbox on server.

Both the wallpaper image (which contains secret mails) and simple mails are downloaded to mail client of user machine. The user will recover the secret mails from the image using a USB token which contains key for the user to decrypt the secret mail message. The user who has the usb token can only view confidential mail message hidden in the form of image. The simple mail messages are as it is downloaded and can be read.

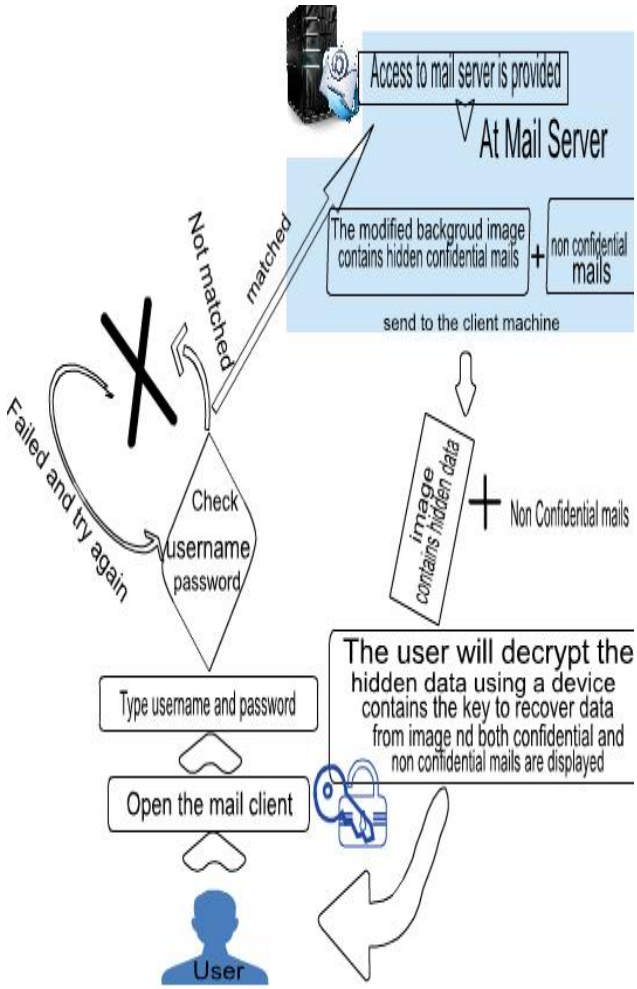

Figure 4. Log in and receiving mail procedure.

3.3 Data Mining

The Data mining is to determine whether the e mail is confidential or not based on some predefined associations rules stored in sender machine. These rules will decide whether the mail sender want to send is confidential and require to be hidden in background image of sender or not. Then if it is confidential EPS is applied on it otherwise it is kept as it is.

3.4 EPS Algorithm

This is the proposed algorithm also known as E-mail Privacy Steganography. This will be a steganographic algorithm used to hide test behind the image and assume image as a 24 bit image means image is large group of pixel and each pixel is of 24 bits. These 24 bits determine color value and has three colors RGB i.e. starting 8 bits are for Red, next 8 bits are for Green and remaining 8 bits are for Blue.

The data is hidden in the LSB of Green and Blue parts only i.e. 2 bits of data can be hidden one in Green and one in Blue. The Red value's Last 2 bits will determine weather data is hidden in this pixel and if hidden then how many bits are hidden. If last 2 bits of red part are 00 then there is no data is hidden in this pixel, if it is 01 then only Blue part contains 1 bit of hidden data i.e. only one bit is hidden in this pixel and if red part's last two bits are 11 then both green and blue parts LSB contains the data.

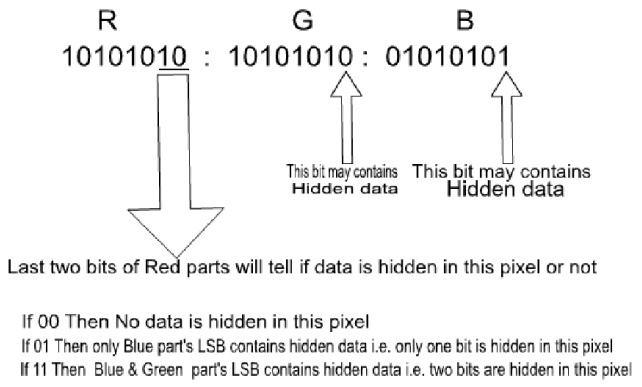

Figure 5. EPS algorithm.

3.4.1 EPS(E-mail Privacy Steganography) Algorithm

First of all consider 24 bit image based on RGB based technique. Start with the pixel according to key K value and select R's value of this Pixel is proceeding as follows:

1. Check last 2 bits of R's value and set them according to our requirement i.e. if we don't want to hide data in this pixel set these bit to 00, if we want to hide only one bit of data in this pixel set these bits to 01 and if we want to hide 2 bits in this pixel set these bits to11.
2. Then if we set last two bits of R's values to 01 hide the one bit of data to be hidden, in the LSB of Blue part.
3. Then if we set last two bits of R's values to 11 hide the one bit of data to be hidden, in the LSB of Blue part and one bit of data in Green's part LSB.
4. Then go to next pixel and repeat the above mentioned procedure.

This algorithm is implemented to hide the data in our proposed system.

4. RESULT

It concludes that our system has been proves a secure model for e-mail communication. In our system we write a mail and the proposed work has been tested using a 1 KB mail message and 500*600 image. The message will hided successfully in the image and no alterations are observed with naked eyes as shown in figure. Hence it is impossible for the intruder to guess that this image take part in communication can have a secret text.

Figure 6. Digital object before steganography.

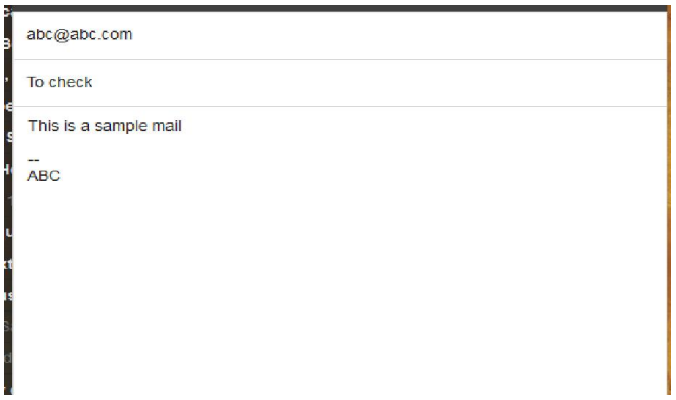

Figure 7. E-mail which will hide within digital object.

Figure 8. Digital object after steganography.

Figure 9. Content of e-mail after unhide.

The our proposed model will be beneficial then the existing PGP approach in some aspects like no compatibility issues, less overhead of keys as key is kept in the usb token and also no doubt less complex as no complex algorithm is used unlike PGP.

5. FUTURE SCOPE & CONCLUSION

The overall conclusion state the system is fully developed to tackle all the problems faced by existing technology PGP for email communication. The proposed system will handle all the aspect of the email security. EPS provide very secure and private mechanism for email.

It can't say this model is fully developed and ready to be used but it can be a better and more secure model for delivering secure e-mails. This can be further used in some areas where security of mails is a crucial factor like defense system. But a lot of work is needed in making it universal email system but it can be a good system if developed.

6. REFERENCES

[1] Pasupuleti Rajesh and Gugulothu Narsimha. 2005. *Privacy Preserving Data Mining by using Implicit Function Theorem*. Vol.5, No.2, March 2013. International Journal of Network Security & Its Applications (IJNSA).

[2] Olivier de Vel. 1999. *Mining Email Authorship*. Defense Science and Technology Organization.

[3] Krishan Gupta, Mukesh Sharma. 2000. *Signature Hiding Standard: Hiding Binary image into RGB based image.* ACM Digital Library. http://dl.acm.org/citation.cfm?id=2677926

[4] Dumitrescu, S., X. Wu and Z. Wang. *Detection of LSB steganography via sample pair analysis*. Springer LNCS, vol.2578, pp.355-372, 2003.

[5] .Wikipedia. *Pretty Good Privacy*. https://en.wikipedia.org/wiki/Pretty_Good_Privacy

[6] C. Kessler. 2001. *Steganography: Hiding Data within Data*. Windows & .NET Magazine. http://www.garykessler.net/library/steganography.html.