An Elucidation on Steganography and Cryptography

Sunita Chaudhary
Dept. of CS & IT
Jagannath University,,Jaipur
er.sunita03@gmail.com

Dr. Meenu Dave
Dept. of CS & IT
Jagannath University,Jaipur
meenu.s.dave@gmail.com

Dr. Amit Sanghi
Dept. of CSE
MEC,Bikaner
dr.amitsanghi@gmail.com

Jaideep Manocha
Dept. of CSE
MEC,Bikaner
a1manocha@gmail.com

ABSTRACT

Safety measures are the main issue while communication over digital networks, but no one can assure delivery of the right contents as burglars may on the way keeping an eye on the communication. To guard the secret data from stealing, various techniques have been implemented to encrypt and decrypt the data. Cryptography and Steganography are the two most famous techniques regarding the same. The first one scrambles the original data and the second one hides the original data under some other media. Both have strong impact when used together instead of using individually. In this paper we are going to discuss the both, while our main focus is on exploring the different techniques of Text Steganography and compare them in terms of robustness and hiding capacity.

CCS Concepts

• **Security and Privacy →Cryptography and Stegneography→Text Stegneography.**

Keywords

Cryptography; Steganography; Encrypt; Decrypt; Conceal.

1. INTRODUCTION

Information travels via computer networks worldwide. The main issue is delivery of the original message. We do not want that our original message to be intercepted by anyone else excluding the recipient so to make our message safe many methods are used. Basically there are three approaches concerning information safety Cryptography, Steganography and their combination [1] as shown in figure 1

1.1 Cryptography

The scrambled writing is called cryptography. In this method the information is visible to everyone but it is written in such a way that the original message cannot be guessed [2]. The scheme is based on encryption, decryption and keys. There is two type of cryptography.

ICTCS '16, March 04-05, 2016, Udaipur, India
© 2016 ACM. ISBN 978-1-4503-3962-9/16/03…$15.00
DOI: http://dx.doi.org/10.1145/2905055.2905249

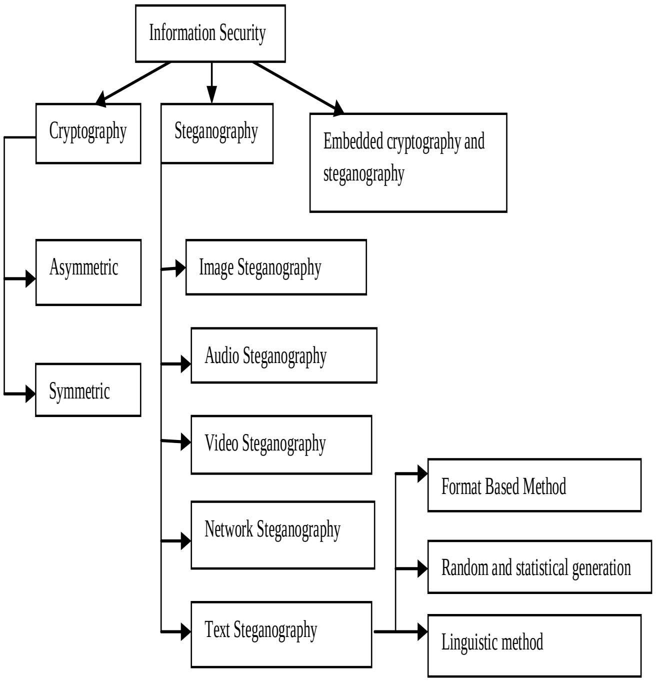

Figure 1. Classification of Information security.

1.1.1 Symmetric key Cryptography

In this method there is a common key shared between both sender and receiver. The same key cannot be used for communication with other party so the drawback of this is that every communication party needs a separate key [2].

1.1.2 Asymmetric Key Cryptography

In this method two keys known as public and private keys are used. Sender encrypts the message by its private /public key and the receiver decrypts the message using the corresponding opposite key [3].

Cryptography scheme is very useful for the following things:

1.2.1 Confidentiality

Information can only be accessed by authorized party and not by anyone else.

1.2.2 Authentication

It can be checked that the information is coming from the right source or not and vice versa.

1.2.3 Integrity

No one can alter the message hence integrity is maintained.

1.2.4 Non repudiation

No one other than the sender and receiver can deny anything regarding the message.

1.2.5 Access Control:

The information can only be accessed by the authenticated parties.

The older methods of cryptography were related to just converting the plain text into cipher text by using substitution method. But the method was so weak and now a day's strong algorithms based on complicated mathematical formulas are in practice like RSA, D-H etc [3]. General form of cryptography is given in figure 2.But on the other hand the transmission of encrypted message may suspect an attacker and the message can be attacked, decrypted or intercepted .To keep the message safer steganography can be used.

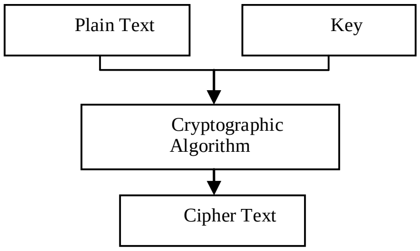

Figure 2. General Form of cryptography.

1.2 Steganography:

Steganography means the art of hiding data and transmission both under other media. It hides the secret data in another file in such a way that only the recipient knows the existence of message [4]. The basic form of transmitting the data is media under media .The multimedia object like audio, video, images are used as a cover sources to hide the data [5]. The basic idea of Steganography is depicted as follows in figure 3.

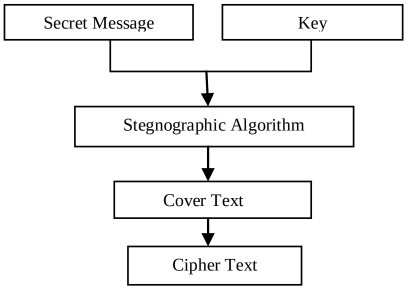

Figure 3.General form of steganography.

Types of Steganography: Text steganography is mainly of five types.

1.2.1. Text Steganography

In this text files are used to hide information. The data can be hidden at the every letter of every word, on every line or on every sentence of text message. Methods available for the same are Format Based, Random and Statistical Method and Linguistics Method.

1.2.2 Image Steganography

In this image works as a cover media .Pixels are used to put the data out of sight. The method is well published and widely used in these days.

1.2.3 Audio Steganography

Audio files are used as cover. This cover can be any auditory format as WAV, AU and MP3 sound documentation. Many approaches for the same are available as Low Bit Encoding, Phase Spread Spectrum.

1.2.4Video Steganography:

Digital video format is used as a cover. It works as carrier of information. Discrete cosine transform is used in this technique. This kind of steganography uses H.264, Mp4, MPEG, AVI formats.

1.2.5 Network or Protocol Steganography

Concealing of information is done using network protocols as CP, UDP, ICMP, IP etc. The Base is OSI layer network model in which covert channels are present those can be used for steganography.

1.3 Embedded Crypto and Stego

Combining Cryptography and Steganography is having a great impact in terms of security .The basic idea behind this is very simple as depicted in the figure 4.

In this method Stenographic encryption is applied first after that cryptography is used. The result is a complicated cipher text. In case of any mishappening with the data the attacker has to do endless efforts thus combining both is more powerful rather than using them individually [5,6].

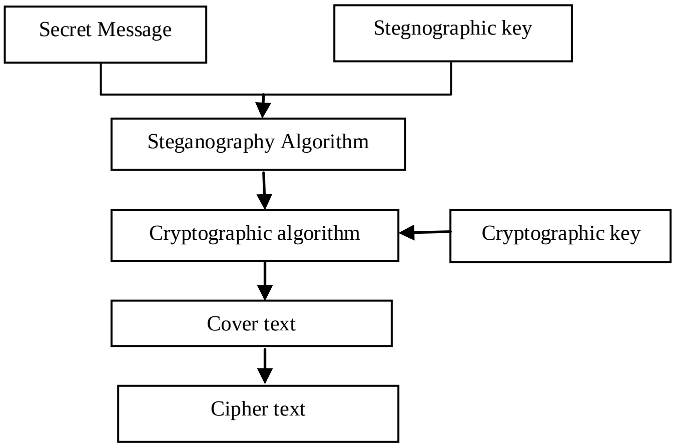

Figure 4. General form of combined cryptography and steganography.

2. TEXT STEGANOGRAPHY

Basically steganography depends upon the cover media i. e depends upon what cover media is used. If the text data is hided under audio it becomes audio steganography and if video is used as a carrier then it becomes video steganography. But hiding text under audio, video or other than any format which is not text is not good due to many technical issues so in practice text under text scheme is used. This is the toughest method than audio, video or network steganography. In this method various styles are used to hide our media as changing the format of text, changing the arrangement of letters, using arbitrary sequence of text, variation in color, size and spelling of the text of sender's interest [7].

The scheme as we said is complicated because in audio and video steganography the image, audio and video have information redundancy that serves as a basis of steganography but in text, data is always indistinguishable so changing the structure of text becomes exigent and most tricky task. On the other hand it has several advantages as it gives easy and speedy communication. It also saves memory as text file consumes minimum bytes as compared to the other media [9, 10]. Available approaches of text steganography are:

2.1 Format based method

Format-based methods use physical format of text to hide secret information. Use of white space, deliberate misspelling, changing color, size or type of fonts, etc are some example of format based text steganography method. Use of white space and deliberate misspelling to hide secret data can be recognized by computer while it is taken as a normal text by human readers. Same way, changing type, color or size of text consider as a normal text by computer but human can immediately recognize this changes[7]. If someone reformat or retype the text then data can be lost. Same way if someone get original document, then by comparison data can be disclosed. Feature coding method and open space methods are example of format based method. The scheme is related with changing the presentation or the format of the text.

We can do various things like, changing color of the text, changing the font of the text, changing, alignment of letters, Changing the white space arrangement in the text, Changing the spelling of the text, Changing text width, Changing length of the text, Adding extra punctuation marks etc.[20]

2.2 Random and statistical generation:

In this method, we are generating random character sequence or random word sequence to use as cover text. This generated sequence can be used as a cover text, to avoid known plain text attack [20]. Random character sequence and random word sequence is generated using statistical properties. The random character sequence must appear random after hiding secret information to everyone, except the ones who are communicating using this. The second approach to character generation is to generate word by using statistical properties of word-length and letter frequency, so words appear like actual words. But because of randomness, both random character sequence and random word sequence are meaningless and that is why it is suspicious. [7, 2, 3].

2.3 Linguistic method

In this category, linguistic structure of text is used to hide secret information. This method considers linguistic properties of generated or modified text to hide secret information. Syntactic structure of text itself can be used to hide secret information [7, 20].

3. EXISTING APPROACHES

Many approaches are developed using the above methods. Here we discuss some of the popular approaches used in current practice.

3.1 Line shift

To hide the original data the line of text is shifted vertically by some degree as 3/10. To hide a 0 line can be shifted upwards and to hide a 1 we shift downwards or vice versa. We can also use no shifting for 0 and shifting up or down for 1.The problem with the method is the information is destroyed in case of retyping the text and also distances of the line shifting can be measured by instruments as by an OCR[10, 12, 13]. The main advantage of shifting algorithm is minimalism in executing, on the contrary hidden data ratio is very less compared to other algorithms of text steganography [13].

3.2 Word Shift

In the scheme the confidential data is kept hidden by placing the words horizontally i. e. by increasing the length of the word in left or right. The method is less noticeable than the line shift because it gives illusion that the text is justified.

But it is also having same problem as line shift i. e information is lost in case of an OCR machine reading or retyping[11,12].Same problem is exist in this method like line shift that is has very low hidden ratio.

3.3 Syntactic method

The method take use of punctuation marks as comma (,), full stop (.) etc. These marks serves as a basis of hiding 0 or 1.The method is very good but requires a lot of care. An intruder having good knowledge of English can intercept because he or she knows that what the exact may position of such marks in a text document and has low hidden ratio [9,10].

3.4 White tag

In this method white space or blank space serves as basis of concealing the information [9]. The method can be used in three different ways:

3.4.1 Inter sentence case

In this scheme space at the end of a sentence is used to hide the secret or confidential data.

3.4.2 *Inter word case*

This method takes use of space available between words.

3.4.3 *End of the line*

This method takes use of the space available at the end of a line.

The difficulty with these schemes is that incorrect use or retyping again makes the hidden data noticeable to an attacker [9].

3.5 Spam text

In the scheme bits are hided in the tags of the markup language file as HTML or XML file. HTML starts and end tags are case insensitive as well as they can occur more than once. Space is also not a considerable thing while writing tags and all of these features can serve as a basis of concealing [13].

3.6 SMS –Texting

People take use of short forms of words while messaging each other. These short forms are called abbreviated words. In this scheme a full word can hide a 1 and the abbreviated word can hide a 0 or vice versa [8].

3.7 Feature Coding

In the method we can alter the features of the text one or more and the altered feature can serve as a basis of steganography. The feature can be style, shape and size of writing a letter or text. As for example size of the dot used in the small English alphabets i and j can be altered to hide a 0 or 1. [13, 14].

3.8 Secret Stenographic code for embedding

The method take use of the article of the English language a, an, to hide 0 and 1[15]. For example to hide a 0 we use article ‘a’ and to hide a 1 we use ‘an’.

3.9 MS word document

In this method some parts of a text documents are relapsed using mimicking and further the relapsed or mimicked parts are used to hide a 0 or 1[17].

3.10 Cricket Score Board

Cricket score board serves as the basis of hiding. As a senseless zero before a number can be used for concealing a 1 and the number without a leading zero can be kept as it is to hide a 0. [18].

3.11 Cascading style sheet

The scheme comes under embedded cryptography and steganography approach discussed above. For the cryptographic part of the scheme RSA is applied and resultant cipher text is embedded with a CSS by using end of line white space scheme explained earlier in this paper [19].

3.12 CASE approach:

This is combination of feature coding method and random character scheme. In this scheme we focus on the capital alphabets of the English language. Shape like horizontal line, vertical line and curve shapes of these alphabets are used as the main feature in this method [20].

4. APPLICATIONS

Steganography is having a wide area of application as Confidential communication, secret data storing and protection of data alteration. It has great importance in the area of Access control system for digital content distribution and Media Database systems also. The uses can differ area by area. Some of the practical practices we can state as:

It can be best way of transmitting news and information which is confidential without having fear of interception.

Steganography is a better solution for accumulation of information on a location. For banking database can be a cover of defense secrets. No one can suspect on it and if it is so no one can extract.

Watermarking can also use steganography methods. Many steganography techniques can accumulate watermarks inside information.

E-commerce is one of the popular areas that can take steganography in to account. In present e-commerce users are authenticated by a username and password but there is no actual verification method of proving that the user is the real card owner.

This kind of verification can be done by using Biometric finger print scanning with unique session IDs covered with the fingerprint images via steganography.

4.1 Cryptography Vs. Steganography

In Cryptography the information is messed up in such a way that it cannot be suspected and if so then cannot be detected. Steganography can be said as layer on information to be sent. This goal of the cover layer is prevention from suspicious.

5. RESULT

Following table 1. represent the comprehensive comparisons of various text Stegnographic approaches in terms of hiding capacity and robustness. In this table it is clear that Feature Coding Method and CASE approach is high in terms of hiding capacity and robustness.

Table 1. Comparison of various text steganographic techniques

S.No.	Method	Criteria	
		Hiding capacity	**Robustness**
1.	Line shift	One Bit per line	Low
2.	Word shift	One Bit per word	Low
3.	Syntactic method	Low	High
4.	White tag	One Bit per space	Low
5.	Spam text	High	Low
6.	SMS texting	High	Low
7.	Feature coding	High	High
8.	SSCE	Low	Low
9.	MS word document	low	High
10.	Cricket Score Board	high	Low
11.	Cascading style sheet	One bit per line/ semicolon	Low
12.	CASE Approach	High	High

In table 2, we can see the CASE approach is very efficient in time overhead and it we can hide all eight bits of one letter of secret message into one letter of cover text at a time, that's why no. of byte hide is same as message text size.

Table 2. Overhead for various message text sizes and time overhead in case approach

	Message Text size (Bytes)	**Cover Text size (Bytes)**	**No. of bytes can hide (Bytes)**	**Time Overhead (ms)**
CASE Approach	400	4000	400	2475
	600	6000	600	2250
	800	8000	800	2375
	1000	10000	1000	2635
	1200	12000	1200	2139

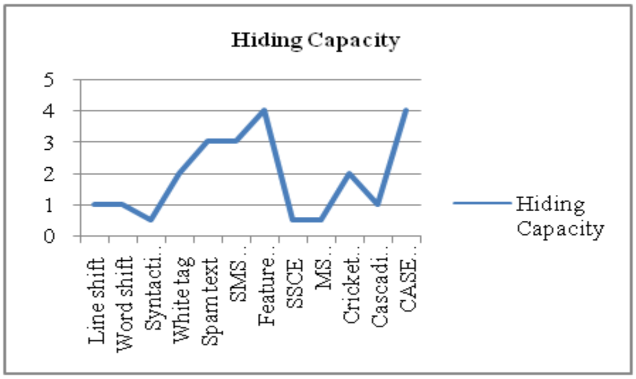

Figure 5. Hiding capacity of various text steganography approaches.

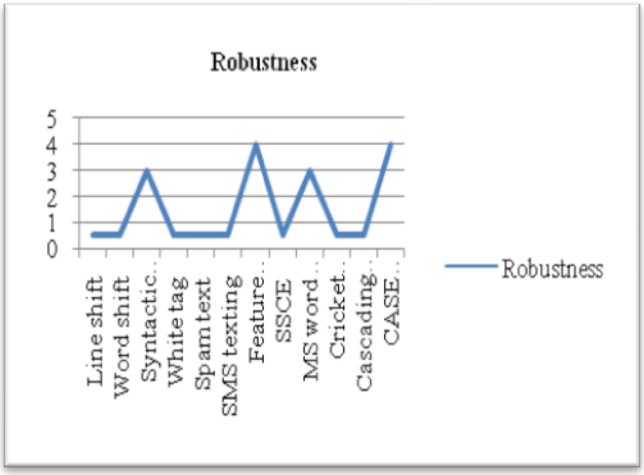

Figure 6. Robustness capacity of various text steganography approaches.

As shown on the above figure 5 & 6 consequently shows hiding capacity and robustness of different text Stegnographic approaches. Line shift and word shift method can hide one bit in a line or word, CSS method can hide one bit per line or after per semicolon and white tag method can hide one bit after every space.

In spite of them Feature coding and CASE method can hide more bit as CASE can hide all eight bits of one letter of secret message into one letter of cover text at a time.

6. CONCLUSION

Behind the interior of Steganography, its uses can either be lawful or illegitimate when in the route of applying the techniques to expose or embed secrets. It is a terrible craze when the techniques are known to scandalous people.

Though, it is a superior thing when it is used for business or humanity as in countrywide security keeping an eye on the criminals and their communication to keep the planet secure. But still, at the same time, when it comes to privacy, the whole scope of steganography can be lost. The technique steganography has its own pros and cons and a large endless area of Implementation.

So finally the thing of our interest in exploring cryptography and steganography in this paper is using them in a combination to achieve the high degree of security. As this is the age of information and muggers are never apart so information security is the prime factor and we are sure that there is a wide possibility of further research in the same area.

Thus in this paper comparative study of various text Stegnographic approaches has been done and the researchers found that feature coding and case approach is more secure and has more hiding capacity.

Further, after analysis of result it has been observed that CASE approach is very robust then other techniques as retyping and reformatting have no effect on hidden message, in the sense that hidden message remains as it is even if we retype the text or reformat the text also robustness is increased due to randomness.

7. REFERENCES

[1]. B. Joonsang, N. Jan, S.-N. Reihaneh, and S. Willy, "A survey of identity-based cryptography," presented at the Australian Unix Users Group Annual Conference, 2004.

[2]. F. Piper and S. Murphy, Cryptography: A Very Short Introduction: Oxford University Press Publishing 2002.

[3]. K. Jonathan and L. Yehuda, Introduction to modern cryptography: principles and protocols: Chapman & Hall/CRC, 2007.

[4]. R. Kefa, "Steganography-the art of hiding data," Information Technology Journal, vol. 3, pp. 245-269, 2004.

[5]. P. Fabien, A. Ross, and K. Markus, "Information hiding-a survey," IEEE,special issue on protection of multimedia content, vol. 87, pp. 1062-1078, 1999.

[6]. R. Joseph and S. Viny, "Cryptography and Steganography-A Survey," International Journal of Computer Technology and Applications, vol. 2, p. 4, 2011.

[7]. K. Benett, “Linguistic steganography- survey, analysis and robustness concerns for hiding information in text,” Purdue University, CERIAS Tech. Report 2004-13, 2004.

[8]. M. S. Shahreza, and M. H. S. Shahreza, “Text steganography in SMS,” 2007 Int. Conf. on Convergence Information Technology, 2007, pp. 2260-2265.

[9]. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” IBM Systems Journal, vol.35, pp. 313-336, 1996.

[10]. M. H. S. Shahreza, and M. S. Shahreza, ”A new approach to Persian/Arabic text steganography,” In Proceedings of 5th IEEE/ACIS Int. Conf. on Computer and Information Science and 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse, 2006, pp. 310-315.

[11]. M. H. S. Shahreza, and M. S. Shahreza, ”A new synonym text steganography,” Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, 2006, pp. 1524-1526.

[12]. S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O. Gorman, “Document marking and identification using both line and word shifting,” INFOCOM’95 Proceedings of the Fourteenth Annual Joint Conf. of the IEEE Computer and Communication Societies, 1995, pp. 853-860.

[13]. J. Cummins, P. Diskin, S. Lau, and R. Parlett, “Steganography and digital watermarking,” School of Computer Science, 2004, pp.1-24.

[14]. J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O. Gorman, “Electronic marking and identification techniques to discourage document copying,” IEEE Journal on Selected Areas in Communication, vol.1, pp. 1495-1504, 1995.

[15]. I. Banerjee, S. Bhattacharyya, and G. Sanyal, “Novel text steganography through special code generation,” Int. Conf. on Systemics, Cybernetics and Informatics, 2011, pp. 298-303.

[16]. S. Bhattacharyya, I. Banerjee, and G. Sanyal, “A novel approach of secure text based steganography model using word mapping method,” Int. Journal of Computer and Information Engineering, vol.4, pp. 96-103, 2010.

[17]. T. Y. Liu, and W. H. Tsai, “A new steganographic method for data hiding in Microsoft word documents by a change

tracking technique," IEEE Transactions on Information Forensics and Security, vol.2, no.1, pp. 24-30, 2007.

[18]. M. Khairullah, "A novel text steganography system in cricket match scorecard," Int. Journal of Computer Applications, vol.21, pp. 43-47, 2011.

[19]. H. Kabetta, B. Y. Dwiandiyanta, and Suyoto, "Information hiding in CSS: a secure scheme text steganography using public key cryptosystem," Int. Journal on Cryptography and Information Security, vol.1, pp. 13-22, 2011.

[20]. sunita chaudhary, p. mathur, T. Kumar, R. Sharma., "A Capital alphabet shape encoding(CASE) based text steganography"on Conference on Advances in Communication and Control Systems 2013 (CAC2S 2013) published by atlantise press, p.p. 120-124.

[21]. Mr. Hitesh Singh, Mr. Anirudra Diwakar, and Ms. Shailja Upadhyaya, "A noval approach to text steganography", Published on IPCSIT vol. 59 (2014) IACSIT Press, Singapore .

[22]. Ms. Latika,Y. Gulati, "A review of text steganography research and development ,"International Journal of Advanced Research in Computer Science and Software Engineering 5(4)April-2015, pp. 871-874.