Different techniques for hiding the text information using text steganography techniques: A survey

V. Lakshman Narayana[1,*], A. Peda gopi[2], N. Ashok Kumar[3]

Vignan's Nirula Institute of Technology and Science for Women, Peda Palakaluru, Guntur, Andhra Pradesh, Guntur.

lakshmanvejendla@gmail.com

ABSTRACT. *Steganography is helping individuals to send private information between two gatherings. It empowers client to conceal information in various advanced mediums. Steganography is of numerous kinds, for example, picture steganography, content steganography, sound/video steganography and so on. Content Steganography is very troublesome than different strategies as a result of less measure of repetition and changes can be distinguished effectively. A portion of the systems of content steganography has been talked alongside attributes and working on it. Sending scrambled messages habitually will draw the consideration of outsiders, i.e. wafers and programmers, maybe making endeavors who break and uncover the first messages. A New approach is proposed in data concealing utilizing between word break up and between passages separating as a cross breed technique. Our strategy offers dynamic created stego-content with six alternatives of most extreme limit as indicated by secrecy message length. This paper additionally broke down the noteworthy disadvantages of each current strategy and how the proposed method resolves such issues.*

RÉSUMÉ. *La stéganographie aide les individus à envoyer des informations privées entre deux assemblées. Il permet au client de dissimuler des informations dans divers moyens avancés. La stéganographie ont plusieurs types, par exemple, la stéganographie sur l'imagerie, la stéganographie par contenu, la stéganographie par son / vidéo, etc. La stéganographie par contenu est très gênante par rapport aux différentes stratégies en raison d'une moindre mesure de répétition et dont les changements peuvent être distingués facilement. Une partie des systèmes de stéganographie par contenu a été mise au point aux côtés d'attributs et y travaille. L'envoi de messages brouillés attirera généralement l'attention des personnes extérieures, comme les wafers et les programmeurs, en faisant peut-être des tentatives qui briseront et dévoileront les premiers messages. Une nouvelle approche est proposée dans la dissimulation de données et utilisée entre la séparation de mots et de passages en tant que technique hybride. Notre stratégie offre un contenu stego créé dynamique avec six alternatives de la limite la plus extrême indiquée par la longueur du message secret. En outre, cet article décrit les inconvénients notables de chaque stratégie actuelle et explique comment la méthode proposée résout ces problèmes.*

KEYWORDS: *steganography, hiding text, text steganography, hiding techniques, randomized techniques.*

MOTS-CLÉS: stéganographie, dissimulation de texte, sstéganographie de texte, techniques de dissimulation, techniques randomisées.

DOI:10.3166/ISI.23.6.115-125 © 2018 Lavoisier

1. Introduction

Data is a critical resource of humankind, whose security is a fundamental concern. Vulnerability increments if taking a shot at constant frameworks which incorporate saving. Odds of assault increments when we transmit information by means of web. A few sorts of assaults are conceivable, for example, listening cautiously, man in the middle assault, phishing assault, dissent of administration and so on.

So to secure our information, we are left with three fundamental arrangements which are by utilizing a private committed channel, cryptography and steganography. Private committed method is tedious and client is limited to a physical point. Cryptography shape the message in some other frame. Couple of cryptography and steganography can likewise be utilized which are known as changeable cryptography (Rani and Chaudhary, 2013).

Data put away is a general term including numerous sub disciplines. A standout amongst the most vital sub disciplines of steganography appeared in Fig.1.

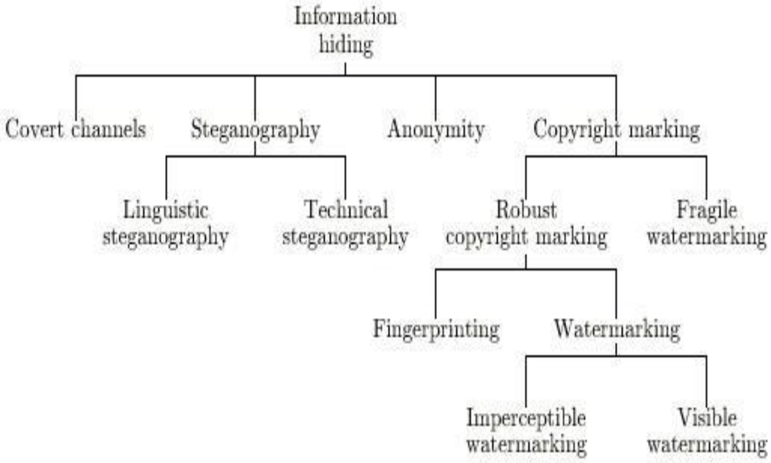


Figure 1. Information hiding techniques classification

Steganography transmits messages with harmless transporter i.e. content, picture, sound and video above correspondence control where the proximity of the message is covered. In Fig.1, steganography is most camouflaging strategies that can organize semantic steganography and fastidious steganography. Etymological steganography depicted as "the quality of utilizing made trademark dialect to cover anchored messages". (Memon *et al.*, 2008; Gutub and Fattani, 2007; Por *et al.*, 2008; Alla *et al.*, 2008; Garg, 2011; Sharma and Kumar, 2013; Kingslin and Kavitha, 2015; Singh *et al.*, 2012; Shahreza, 2005)

Steganography is the craft of concealing information inside any advanced medium like sound, picture, video, content, convention and so on (Gupta and Gupta, 2011). Frequent terms utilized in steganography are:

Cover Object: Text, sound, video, picture utilized for inserting information is known as cover problem. It is otherwise called craft question.

Secret information/message: The information which is to be inserted in a cover question is known as secret message.

Stego Object: It is the resultant yield acquired in the wake of installing which is known as stego question.

While outlining a safe steganographic framework, following focuses were considered: (I) System is single, remain solitary should meet client prerequisite of classification, genuineness and honesty (ii) User ought to be given by an incognito channel to conceal secret correspondence, (iii) Based on the specialized and physical necessities client ought to have the capacity to have an adjusted parameter determination choice (Dhanani and Panchal, 2013). There are four crucial sort of Steganography 1. Content - Steganography 2. Picture - Steganography 3. Sound - Steganography 4. Video – Steganography.

2. Literature survey

Brassil *et al.* (2004) gave the fundamental idea of report coding procedures in his paper by proposing life-move coding, word-move coding and feature coding (character coding) to disable unlawful scrambling of record scattered by PC make. Line-move coding is a framework for changing a record by vertically moving the areas of substance lines to especially encode the chronicle. Word-move coding is a structure to change a report by on a level plane moving the areas of words inside substance lines to strikingly encode the record.

Wayner (2013) proposed an impersonation calculation to make repeated substance that is apparently similar to the authentic structure of the essential substance. The creator utilized an approach of syntactic standards to settle on stegotext and the decision of each word picks how mystery message bits are encoded. The etymological essentials depend upon static sentence structure which surmises the vernacular structures must be outlined out before the estimation can be utilized. Undeniably, the estimation produces setting free structures. The framework's client must game plan a speech structure that he wishes the substance to imitate.

3. Diverse techniques of text - steganography

Content steganography can be extensively gathered into three sorts:

(1) Association based Random method

Association based techniques (Popa, 1998) consolidate changing physically the setup or design of substance to hide the data. This method has certain imperfections. On the off chance that the stego report uses word processor to open data, mixed spellings what's inexorably, extra void field will get perceived. Misused printed styles can empower vulnerability. Additionally important data is used, separating this plaintext and the suspected steganographic substance makes forbidden parts of the data amazingly noticeable (Aarwal, 2013).

(2) Etymological methods

Phonetic steganography particularly contemplates the etymological properties with changed substance, and everything considered, utilizes semantic structure where messages are hidden. CFG impact is utilized for covering the bits which addresses of left branch is "0" and right branch relates to '1'. This procedure has a few disadvantages. Beginning with a little sentence, structure will provoke some portion of substance accentuation. Moreover, notwithstanding how the substance is phonetically faultless with non appearance.

3.1. Cryptography

Cryptography is connected with the route toward changing over standard plain substance into incomprehensible substance and the different way. It is a methodology for securing and transmitting data in a particular shape with the objective that those for whom it is proposed can read and process it. Cryptography shields data from thievery or change, and in addition be used for customer affirmation.

Earlier cryptography was feasibly synonymous with encryption anyway nowadays cryptography is fundamentally established on logical speculation and programming building practice.

Present day cryptography stresses with:

Security - Information can't be understood by anyone

Respectability - Information can't be balanced.

Non-repudiation - Sender can't deny his/her points in the transmission of the information at a later stage

Affirmation - Sender and beneficiary can confirm each

Cryptography is used in various applications like setting aside extra cash trades cards, PC passwords, and web business trades.

Three sorts of cryptographic strategies used when all is said in done.

- Symmetric-key cryptography
- Hash limits.
- Open key cryptography

3.2. Image steganography

As the name proposes, Image Steganography alludes to the way toward concealing information inside a picture document. The picture chose for this reason for existing is known as the cover-picture and the picture got after steganography is known as the stego-picture.

How is it done?

A picture is spoken to as a N*M (in the event of greyscale pictures) or N*M*3 (in the event of shading pictures) lattice in memory, with every section speaking to the power estimation of a pixel. In picture steganography, a message is inserted into a picture by modifying the estimations of a few pixels, which are picked by an encryption calculation. The beneficiary of the picture must know about a similar calculation so as to known which pixels he or she should choose to separate the message.

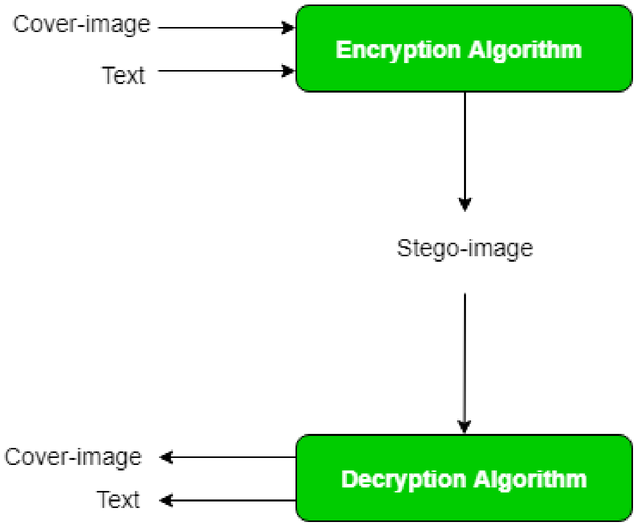


Figure 2. Process of image steganography

Recognition of the message inside the cover-picture is finished by the procedure of steganalysis. This should be possible through examination with the cover picture, histogram plotting, or by commotion recognition. While endeavors are being put resources into growing new calculations with a more prominent level of resistance against such assaults, endeavors are likewise being dedicated towards enhancing existing calculations for steganalysis, to identify trade of secret data between psychological oppressors or criminal components.

An image histogram H(s) center of pixels are calculated as

$$C(h(x)) = \frac{\sum_{x=0}^{n} xh(x)}{\sum_{x=0}^{n} h(x)}$$

Where n is the min and x is the max pixel count.

3.3. Text steganography

Content steganography is a process which hides a secret data behind other content. Text Aberration of any character M(N) from its neighborhood ℓ(M(N)) as far as Standard Deviation of ℓ(M(N)) is specified as δ (M(N), ℓ(M(N))). It is a quantifier that gives the possibility of the measure of deviation of the character from its word.

$$\mathbf{M(N)} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j) - J(i,j))^2$$

(1) Line shift

In this method the secret data is hidden inside another line with a specific degree. The covered data has 2 lines in either directions of secret data. The hidden data bits are marked as 0 and the lines are represented with 1 and the cover text is also represented with 1. The angle position is checked by the center position of the lines arranged and the secure data can be accessed.

(2) Word shift

In this technique, secret data is masked on a plain text by randomly choosing 0 and 1 for every character. Middle square method is used for arranging the data in different portions. Nevertheless, in case somebody recognize the count of portions, he can isolate the stego contented and the count and get the anchored substance by using the ability.

(3) Word mapping

This strategy scrambles a problem message using obtained official cream and a while later displays the ensuing figure content, accepting two bits as quickly as time permits. The embeddings locations are stored and sent to the gatherer close to the stego question.

(4) Proposed technique

In this paper we will propose the accompanying four sections.

1. Enrollment.

2. Encode content data.

3. Mapping through that XML construction.

4. Unscramble that data to another dialect of condition.

Enlistment: Registration is a strategy for authoritatively recording something. Typically something is enrolled to assert more rights, or to ensure possession, or in light of the fact that the law says it must be enlisted to be utilized lawfully. An enroll was an expansive book. It was utilized like a journal to record business dealings or different occasions. To enter individual data about that specific client. To enter that username and secret key distinguishing proof instrument. To give that effective arrangement distinguishing proof data. To get that window in the shape message change. Scramble Text data: To give that any message data that can be changed over in the shape bit organize ID way. To create bit arrange recognizable proof process can be associated with the frame substitution strategy. To substitution system is called as Mapping through that XML dialect to create one of word reference minute distinguishing proof process. Mapping through that XML outline: To give that full portrayal of data through that arrangement procedure to present in the frame tree structure distinguishing proof system. After that to produce ID of right basic leadership ID process. To uncover information data right now accessible information. Decode that data to another dialect of condition: After encoding that data it very well may be exchanged and put away inside a similar view message. This view message recognizable proof process is called as new create XML mapping which can be produced inside that database.

Flow Diagram of the Model

Another strategy is introduced utilizing white/invalid space system for concealing the secret message in content steganography.

In this method a solo space is translated as "0" though 2 spaces are deciphered as "1". The disservice of this procedure is it uses a lot of room to encode couple of bits. For instance, a character is likeness 8 bits, and it requires around 8 interspaces to instruct one character (Low *et al.*, 1995).

For concealing secret message inside a cover message, another method on content steganography with white/invalid space strategy.

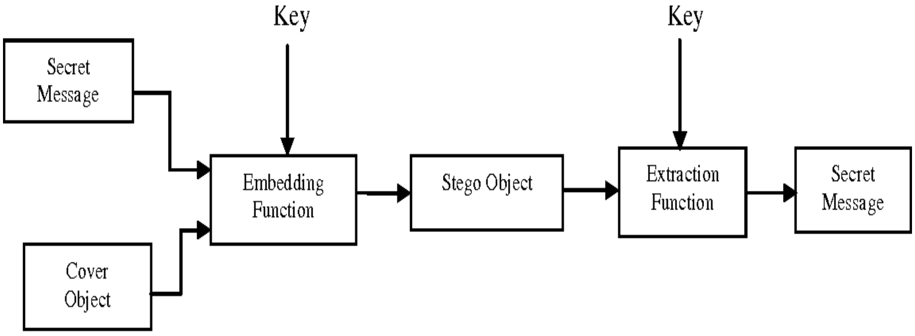


Figure 3. Utilizing white/invalid space system for concealing the secret message

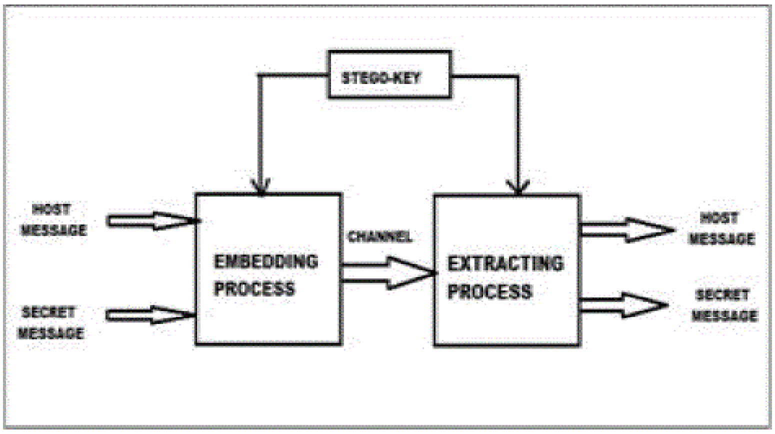
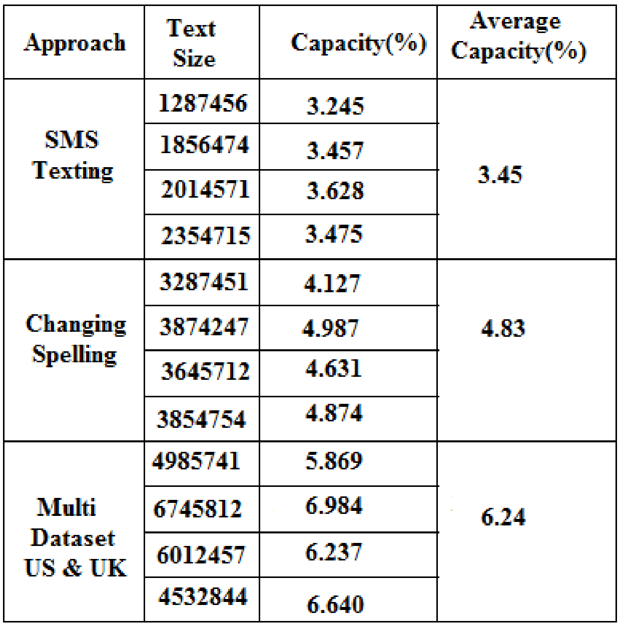


Figure 4. Embedding Text using stego key

The Method introduced is contrasted with the existing methods and the comparative outcomes are depicted in below table.

Table 1. Assessment of text hiding rate



Approach	Text Size	Capacity(%)	Average Capacity(%)
SMS Texting	1287456	3.245	3.45
	1856474	3.457	
	2014571	3.628	
	2354715	3.475	
Changing Spelling	3287451	4.127	4.83
	3874247	4.987	
	3645712	4.631	
	3854754	4.874	
Multi Dataset US & UK	4985741	5.869	6.24
	6745812	6.984	
	6012457	6.237	
	4532844	6.640	

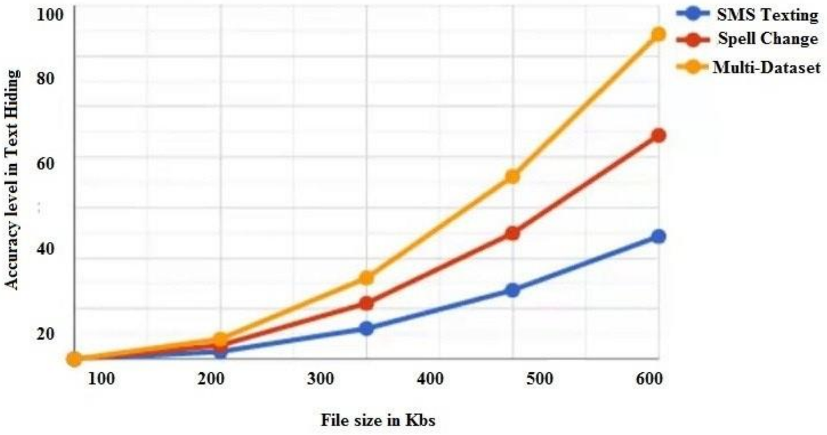


Figure 5. Accuracy level in text hiding methods comparison

5. Conclusion

A few research work is done in the territory of content steganography. With the progression of innovation and instruments accessible, it is currently basic to build up some steganography calculation which can withstand against the attacks. We have introduced another approach of content steganography strategy utilizing between word and interparagraph dispersing for concealing data. This is on account of the shrouded information will be annihilated once the spaces are erased by some word handling programming. Other than that, it is critical to enhance the limit of the implanted plan by mulling over other pressure strategy. The future work ought to be engaged towards streamlining the power of the interpreting calculation. This is on the grounds that the shrouded information will be devastated once the spaces are erased by some word handling programming. Other than that, it is imperative to enhance the limit of the inserted conspire by thinking about other pressure strategy.

References

Agarwal M. (2013). Text steganographic approaches: A comparison. *International Journal of Network Security & Its Applications (IJNSA)*, Vol. 5, No. 1, pp. 91-106. https://doi.org/10.5121/ijnsa.2013.5107

Ali A. A. (2013). New Text Steganography Technique by using Mixed-Case Font. *International Journal of Computer Applications*, Vol. 62, No. 3, pp. 6-9.

Alla K., Prasad S. R., Ram S. (2008). A novel Hindi text steganography using letter diacritics and its compound words. *International Journal of Computer Science and Network Security*, Vol. 8, No. 12, pp. 404-409.

Bennett K. (2004). Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text. *Citeseer*. https://doi.org/10.1.1.158.8602

Dhanani C., Panchal K. (2013). Steganography using web documents as a carrier: A survey. *International Journal of Engineering Development and Research*. https://doi.org/10.11241/ijedr.21412

Garg M. (2011). A novel text steganography technique based on html documents. *International Journal of Advanced Science and Technology*, Vol. 35, pp. 129-138.

Gupta S., Gupta D. (2011). Text-steganography: Review study & comparative analysis. *International Journal of Computer Science and Information Technologies*, Vol. 2, No. 5. https://doi.org/10.21421/ijcsit.11421

Gutub A., Fattani M. (2007). A Novel Arabic Text Steganography Method Using Letter Points and Extensions.

Kingslin S., Kavitha N. (2015). Evaluative approach towards text steganographic techniques. *Indian Journal of Science and Technology*, Vol. 8, No. 29. https://doi.org/10.17485/ijst/2015/v8i29/84415

Low S. H., Maxemchuk N. F., Brassil J. T., O'Gorman L. (1995). Document marking and identification using both line and word shifting. *Proceedings of INFOCOM'95*. https://doi.org/10.1109/INFCOM.1995.515956

Memon J. A., Khowaja K., Kazi H. (2008). Evaluation of steganography for Urdu/Arabic text. *Journal of Theoretical and Applied Information Technology*, pp. 232-237.

Popa R. (1998). An analysis of steganographic techniques. *The Politehnica University of Timisoara, Faculty of Automatics and Computers*.

Por L. Y., Ang T. F., Delina B. (2008). WhiteSteg- a new scheme in information hiding using text steganography. *Transactions on Computers,* Vol. 7, No. 6, pp. 735-745.

Rani N., Chaudhary J. (2013). Text steganography techniques: A review. *International Journal of Engineering Trends and Technology*, Vol. 4, No. 7, pp. 3013-3015.

Shahreza M. S. (2005). A new method for steganography in HTML files. *Advances in Computer, Information, and Systems Sciences, and Engineering,* pp. 247-252. https://doi.org/10.1007/1-4020-5261-8_39

Sharma V., Kumar S. (2013). A new approach to hide text in images using steganography. *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, No. 4, pp. 701-708.

Singh P., Chaudhary R., Agarwal A. (2012). A novel approach of text steganography based on null spaces. *IOSR Journal of Computer Engineering*, Vol. 3, No. 4, pp. 11-17.