

An Antonym Substitution-based Model on Linguistic Steganography Method

Fawwaz Zamir Mansor¹, Aida Mustapha², Roshidi Din³, Azizi Abas⁴, Sunariya Utama⁵

^{1,2}Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja 86400, Batu Pahat, Johor, Malaysia

^{3,4,5}School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Article Info

Article history:

Received May 1, 2018

Revised Jul 10, 2018

Accepted Jul 25, 2018

Keywords:

Antonym Substitution

Linguistic semantic

Synonym substitution

ABSTRACT

This paper implement the antonym procedure in substitution-based approach for linguistic steganography that is expected to be an alternative to the existing substitution approach that using synonym. This method is hidden the message as existing approach, however that is changing the semantic of the stego text from cover text as the medium. This paper proposed the development method is named Antonym Substitution-based that applied based on implementation linguistic steganography in covering the secret message. This approach refines the word choices makes antonym substitution-based approach linguistically is comparable with the previous effort of synonym substitution. Furthermore, this paper also used the verification and validation approaches as the procedure to evaluate performance of this method. The evaluation through verification approach used in order to determine the effectiveness and quality performance. This proposed method is expected has become the new technique implementation of linguistic steganography in process securing secret message that hide in the text medium.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Fawwaz Zamir Mansor,
Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia,
Parit Raja 86400, Batu Pahat, Johor, Malaysia.
Email: aidam@uthm.edu.my

1. INTRODUCTION

The implementation text document is one of most important medium in media information such letter appointment. The wishes of text documents as the important point of information are still high in the domain of business and academic. Meanwhile, the most other medium information such as image, audio and video mostly implemented in entertainment environment Therefore, text documents should be a concern for all of people due to the value of information it carries and the high risk of the communication channel [1]. The irresponsible third party perhaps irregularly tempered the information to for their purpose. One of solution to deal that issue is implementation of information hiding area named steganography.

Steganography as the category information hiding that efforts to hide the hidden messages into innocuous looking media in a camouflage manner which avoids suspicions from unintended recipients. There are lot of and variable types of media in which secret messages can be embedded such as images, text-based documents, video and/or audio files. The implementation of text steganography is hiding the hidden message in the text medium that the third party is unable to discover the existences of message in text. Its means, the implementation steganography in medium of expectedly to make the hidden information invisible and could be noticed by unauthorized party. The implementation of steganography itself is divided into two categories. Firstly; the implementation of steganography in different mediums of image, audio, video and other digitally invisible code namely technical steganography [2]. Secondly; the implementation of steganography in the

medium of text is known as text steganography. It is anticipated that the steganography approach can give some solution for safeguarding the security of information in the text media [3]. The implementation of text steganography is by hiding the secret text in the medium of text so that the third party is unable to discover the existence of the message. In other words, steganography in the medium of text can make the secret information invisible and unnoticed for third party to see or detect, and it is directed to the appropriate receivers to apprehend the information. However, steganography in the text domain is the most challenging method of operation compared to the other domains. The challenge in implementing steganography in the text domain since the text file has a small quantity to hide information [2].

There are two groups of text steganography which are linguistic and format-based steganography. Linguistic steganography is dependable on the linguistic order of the sentence of text. The development of format-based steganography consists of two kinds of techniques, which are word-rule based and feature-based. Word-rule based is the technique that embeds the hidden message based on word pattern by shifting in the text. The techniques consist of line-shift coding and word-shift coding. Line-shift coding hides the hidden message with vertically shifting hidden message in text lines. Meanwhile, word-shift coding hides the hidden message with horizontally shifting the hidden message in length between words [2]. The second type format-based steganography is feature-based which can be defined as a technique that alter unique feature characteristic in text based on code words. This technique covers hidden message based on pattern letter or length of the word that conceals so no changes happen in the text [3]. In linguistic steganography, the level of security in secret message is low in making it prone to attack. The implementation of substitution-based approach in linguistic steganography is using the synonyms. However, the literature has shown that the use of synonym in substitution leads to a stego message that is easy to guess [4]. The detectors can discover the hidden messages exist in the text, if they notice the changes in the analyzed text [5]. Low security is also the issue in linguistic steganography approach especially in synonym-based substitution [6]. One evidence is the synonym paraphrasing technique in Spanish language that is easily to attack [7].

To consider for this issue, this paper proposes for the use of antonym substitution-based, in effort to mask the meaning of the original message. Most existing study of substitution-based linguistic steganography apply synonym substitution-based approach, therefore, there is a need to study antonym-based approach as an alternative to implement in linguistic steganography. The effectiveness of this study will be evaluated using verification and validation.

2. RELATED WORKS

The existence of intruders in the communication technology enables anyone to retrieve information easily. Irresponsible intruders or attackers may disclose the secret information to uninvolved parties to check or modify it for abusing that information [8]. The impact of the easier access on the internet making of the increasing of the chance of attackers and intruders, it is a responsibility to take additional measures to ensure the right is well protected. One of the solutions in dealing this issue is to secure the circumstance by using steganography, which is particular linguistic steganography in this paper. This paper elaborated the several development and function in steganography especially in linguistic steganography that begin with the linguistic semantic reviews.

2.1. Linguistic Semantic

Linguistic semantic is the study of significance that is used by humans to express about themselves through language [9]. It is often used in ordinary language to denote a problem of understanding that comes down to word selection. In English language for example, the synonym of the word "presence" is "attend", but its antonym is "absence". One can indicate the two different meanings based on the definition for synonym and antonym.

A synonym is a word that has the same, or almost the same, meaning as another word. Synonyms and antonyms are used by teachers, students, writers, editors, poets, and songwriters. Synonyms are usually used to define a word with the same meaning but presented in another word. The difference with antonym is that it is used to define a word that is against its meaning and is absolutely a different word.

Meanwhile, antonym change the word in terms of giving an opposite meaning from the actual word. In Table 1 is showed some example of dictionary several words with their relevant antonym which is showed the word that has antonym that most possible used word encode for hiding secret message in the text.

Table 1. Substituted-based using Antonyms Words

Word	Antonym	Word	Antonym
absence	<i>presence</i>	approval	<i>disapproval</i>
accept	<i>refuse</i>	approached	<i>receded</i>
accurate	<i>inaccurate</i>	abundant	<i>scarce</i>
advantage	<i>disadvantage</i>	advance	<i>retreat</i>
alive	<i>dead</i>	artificial	<i>natural</i>
always	<i>never</i>	ascend	<i>descend</i>
ancient	<i>modern</i>	attack	<i>defence</i>
answer	<i>question</i>		

2.2. Linguistic Steganography Method

Linguistic steganography is altered the hidden secret message that information concerning the procedure of words and linguistic modification when covering the hidden message. One general technique in linguistic steganography that usually implement is synonym substitution method. The synonym substitution-based able to implement in several languages which the assumption the language the text is of has synonym word. Some of previous researchers effort of linguistic steganography field within the last decade is shown in Figure 1 as follows.

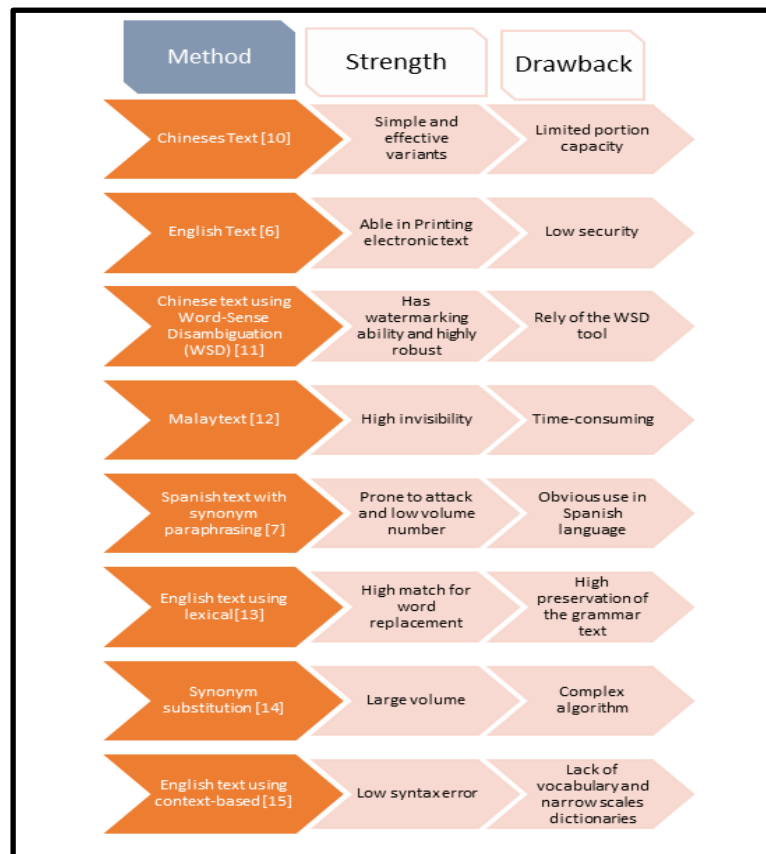


Figure 1. Researcher efforts in linguistic steganography method

Based on Figure 1, it is found that most implementation of steganography is focusing on format-based steganography as opposed to linguistic steganography. Within the area of substitution-based approach in linguistic steganography, only the use of synonyms has been investigated in the past research. However, synonym substitution has expected to improve the cover text that generated with same meaning as plain text. This paper elaborate the mechanism of antonyms in substitution-based approach as the alternative to the standing synonym-based approach.

3. ANTONYM-BASED SUBSTITUTION

This paper discuss a method that used as tool of hiding message in a cover text using the antonyms words from the Thesaurus of Antonyms database. The tool that is developed is called Antonym Substitution-based (ASb) Steganographic Tool based on linguistic steganography [16]. This approach maintains the syntactic and semantic structure of the cover text such that it appears harmless to any unintended recipient. The encryption process is based on word replacement, with words being replaced by their word absolute antonym. It begins by feeding the system with cover text, wordlist and secret text. Then, the system will read the cover text, word by word simultaneously. If the word matches in wordlist found (bit=1), the word will be replaced by its antonym. Once it completes reading the cover text, the stego text will be generated.

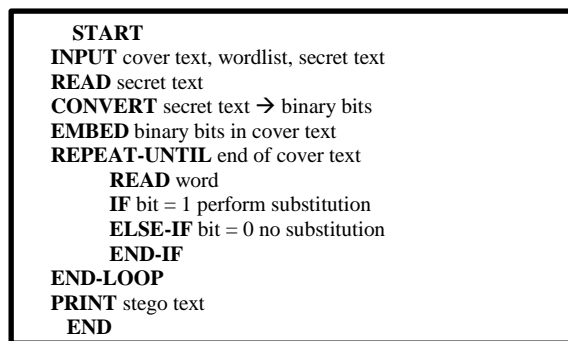


Figure 2. Pseudo code ASb model

From Figure 2, it is shown that the pseudo code as the logical design. The implementation of flow process and pseudo code begins the secret text is processed in a substitution environment where ASb it be applied into secret text. The secret text will them be converted to binary bits and embedded into cover text, which in turn will be substituted with antonym. Then, the antonym text will be taken from antonym wordlist and replaced the identified text in order to generate a stego message. At last, processing decode the stego text, the antonym text that was applied will be extracted, and hence the secret text will be displayed. In term of logical design, the process model of ASb is showed in Figure 3 as follows.

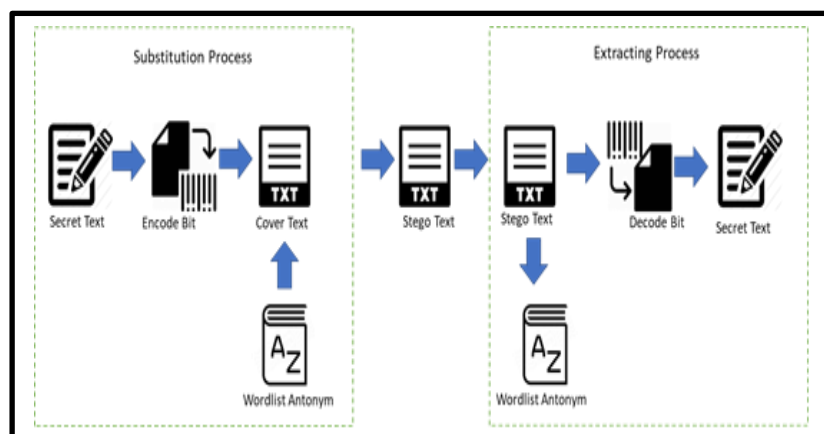


Figure 3. Substituted-based Steganography Model using Antonym

Figure 3 shows the proposed approach has been applied into a tool called the ASb Steganographic Tool, where the algorithm of the proposed approach is implemented into this tool. The interface of the tool established using GUI of VB.NET programming language. It also shows the user interface that is concerned with how users add information to the tool and how the tool presents information back to them interface of ASb. The main characteristic of ASb is that the tool provides spaces for user to place the secret text. The, ASb system approach is tested with two categories plain texts that are cover text which select Reuter News

21578 and secret text which select sentence in the language. The classifications of input datasets are compared based on length of characters of the text and capacity size bit in the text [17, 18]. Every set of input data set was entered into ASb tool and the stego text generated were recorded.

4. EVALUATION RESULTS

This section discusses about the analysis and result performance of proposed study based on the verification evaluation. Table 2 shows the length character in stego text for antonym and synonym stego text. It shows the result of number character of stego text 1 (ST 1) until stego text 20 (ST 20). Based on Table 2, there are changes of character for each stego text from cover text. Its mean all of binary bits are embedded in obtained the stego text.

Table 2. List of characters length of stego text toward cover text

No.	Length of Cover Text (Characters)	Length of Antonym Stego Text (Characters)	Length of Synonym Stego Text (Characters)
1	32834	32896	32898
2	33020	33082	33084
3	31525	31525	31520
4	31156	31588	31584
5	32622	32687	32676
6	34765	34831	34826
7	36772	36845	36835
8	33903	33990	33986
9	30117	30177	30189
10	32617	32682	32678
11	33225	33291	33287
12	32323	32384	32386
13	31132	31132	31126
14	33525	31194	31194
15	34365	34433	34442
16	36522	36621	36627
17	31124	31183	31187
18	30129	30186	30178
19	34141	34205	34204
20	31198	31257	31262

The input datasets in cover text are investigated using 20 different types of length character as following respectively. The size bits are based one capacity of size bitthe datasets. Similarly, the size bit consists of 20 secret texts, which are ordered as follows; 33666 bytes, 34236 bytes, 32234 bytes, 32137 bytes, 31326 bytes, 33874 bytes, 35987 bytes, 32968 bytes, 39869 bytes, 31425 bytes, 32378 bytes, 31252 bytes, 30114 bytes, 32465 bytes, 33785 bytes, 35485 bytes, 30207 bytes, 29856 bytes, 33646 bytes, 30453 bytes. It also involved 20 secret texts with capacity size bit ordering in the range between 36 bytes and 38 bytes. The relationship of input datasets between secret text and cover text based on size bit is shown in Table 3.

Table 3. List of size bit of stego text and secret text

No.	Size bit Cover Text (bytes)	Size bit Secret Text (bytes)	No.	Size bit Cover Text (bytes)	Size bit Secret Text (bytes)
1	33666	37	11	32378	37
2	34236	38	12	31252	37
3	32234	37	13	30114	37
4	32137	36	14	32465	37
5	31326	37	15	33785	37
6	33874	37	16	35485	37
7	35987	37	17	30207	36
8	32968	37	18	29856	37
9	29869	37	19	33646	38
10	31425	38	20	30453	37

The result from experiment in this section showed the size bit of stego text that had been embedded with secret text. The first approach that measure the size bit is used antonym, followed by synonym. The number of amount of size bit stego text using both approaches is shown in Table 4 as follows.

Table 4. List of size bit of stego text toward secret text

No.	Size bit Antonym Stego Text (bytes)	Size bit Synonym Stego Text (bytes)	Size bit Secret Text (bytes)
1	33730	33732	37
2	34301	34338	38
3	32843	32839	37
4	32741	32733	36
5	31918	31920	37
6	33938	33970	37
7	36055	36079	37
8	32968	32965	37
9	29925	29975	37
10	31484	31481	38
11	32439	32476	37
12	31311	31352	37
13	30171	30174	37
14	32526	32565	37
15	33849	33885	37
16	35913	35985	37
17	30264	30207	36
18	29912	29956	37
19	33709	33746	38
20	30510	30553	37

Table 4 shows the size bit stego text of antonym and synonym toward the secret text. It showed the larger size bit of secret text influenced generated larger stego text size bit capacity. Both antonym and synonym stego text output show the size of stego text generated were almost same for each dataset tested. This mean that the antonym was comparable with the synonym

Next, the experiment is about size bit stego text toward cover text is similar with size bit stego text toward secret text. However, this experiment compared the size bit between cover text and stego text. It similar with the previous experiment that is discovered number of stego text. It also used the antonym and synonym approach. This result of experiment used in order to compare the number of size bit stego text that had been discovered with the original capacity of size bit the cover text. The approach that measures the size bit is antonym and synonym that comparing the number of stego text and cover text obtained which is showed in Table 5 as follows.

Table 5. List of size bit of stego text toward cover text

No.	Size bit Cover Text (bytes)	Size bit Antonym Stego Text (bytes)	Size bit Synonym Stego Text (bytes)
1	33666	33730	33732
2	34236	34301	34338
3	32234	32843	32839
4	32137	32741	32733
5	31326	31918	31920
6	33874	33938	33970
7	35987	36055	36079
8	32968	32968	32965
9	29869	29925	29975
10	31425	31484	31481
11	33666	32439	32476
12	34236	31311	31352
13	32234	30171	30174
14	32137	32526	32565
15	31326	33849	33885
16	33874	35913	35985
17	35987	30264	30207
18	32968	29912	29956
19	29869	33709	33746
20	31425	30510	30553

From Table 5, the size bit stego text of cover text in both approach increased. All the stego text that generate from antonym and synonym approach shows the increment of 0.18% to 0.20% of size of bytes. The size of both approaches also comparable each other.

5. CONCLUSION

This paper is elaborated development the new linguistic steganography approach named antonym substitution. This approach obtain the idea implementation of substitution process hiding that targeted the word that has antonym. . A tool has been developed to test the proposed approach and it has been verified. This proposed method has been verified based on its character length stego text towards the cover text, bit size types of the secret text towards the stego text and bit size types of the cover text towards the stego text. For future work, the ASb is expected continue in producing innocuous text and increase utilization cover texts.

ACKNOWLEDEMENT

The authors wish to thank the Ministry of Higher Education Malaysia (MoHE) in funding this study under the Fundamental Research Grant Scheme (FRGS), SO Code 13576 and Research and Innovation Management Centre, Universiti Utara Malaysia, Kedah for the administration of this study.

REFERENCES

- [1] P. Sheldon. *Is Medium the Message? Perceptions of and Reactions to Emergency Alert Communications on College Campuses*. In *Digital Transformation in Journalism and News Media*. 2017: 467–479.
- [2] D. Bharti, A. Kumar. *Enhanced Steganography Algorithm to Improve Security by using Vigenere Encryption and First Component Alteration Technique*, *Int. J. Eng. Trends Technolog*. 2014; 13(5):242-246.
- [3] M. V. Nasab, B. M. Shafiei. *Steganography in Programming*, *Aust. J. Basic Appl. Sci*. 2011: 3(12):1496-1499.
- [4] S. Bhattacharyya, A. P. Kshitij, G. Sanyal, A Novel Approach to Develop a Secure Image based Steganographic Model using Integer Wavelet Transform, in *Proceedings of the International Conference on Recent Trends in Information, Telecommunication and Computing (ITC)*, 2010; 173-178,
- [5] S. Dulera, D. Jinwala, A. Dasgupta. Experimenting with the Novel Approaches in Text Steganography, *Int. J. Netw. Secur. Its Appl*. 2011; 3(6), 213-225.
- [6] M. H. Shirali-Shahreza and M. Shirali-Shahreza, A New Synonym Text Steganography, *Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 2008; 1524-1526.
- [7] A. Munoz, J. Carracedo, I. A. Alvarez, Measuring the security of linguistic steganography in Spanish based on synonymous paraphrasing with WSD, in *Proceedings of the 10th IEEE International Conference on Computer and Information Technology*, 2010; 965-970.
- [8] M. M. Amin, M. Salleh, S. Ibrahim, M. R. Katmin and M. Z. I. Shamsuddin, Information Hiding using Steganography, in *Proceedings of the 4th National Conference on Telecommunication Technology*. 2003; 21–25.
- [9] F. Wang, L. Huang, Z. Chen, W. Yang H. Miao. A Novel Text Steganography by Context-Based Equivalent Substitution, 2013.
- [10] L. Yuling, S. Xingming, G. Can and W. Hong, An Efficient Linguistic Steganography for Chinese Text, in *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*, 2007; 2094-2097.
- [11] H. Lu, L. JianBin, F. TianZhi DingYi, An anti-attack watermarking based on synonym substitution for Chinese text, in *Proceedings of the 5th International Conference on Information Assurance and Security*, 2009; 356-359.
- [12] H. Z. Muhammad, A. Shakil, S.M.S.A.A. Rahman. Synonym-based Malay linguistic text steganography, in *Proceedings of the 2009 Innovative Technologies in Intelligent Systems and Industrial Applications*. 2009; 423-427.
- [13] C.-Y. Chang and S. Clark, Practical Linguistic Steganography using Contextual Synonym Substitution and Vertex Colour Coding, *Emmlp*.2010; 40(2):403-448
- [14] C. Qi, S. Xingming and X. Lingyun, “A secure text steganography based on synonym substitution”, in *Proceedings of the IEEE Conference Anthology*, 2013:1–3.

-
- [15] F. Wang, L. Huang, Z. Chen, W. Yang, and H. Miao. *A Novel Text Steganography by Context-Based Equivalent Substitution*. Retrieved from <http://or.nsf.gov.cn/handle/00001903-5/414117>, 2013.
- [16] R. Din, A. Samsudin. *Intelegent steganalytic system: Appication on natural lanugae environment*. *WSEAS transaction on System and control*. 2009; 4(8): 379-388.
- [17] R. Din, Z.C. Ani, A. Samsudin. A Formulation of conditional states on steganaysis approach. *WSEAS transaction on mathematic*. 2012; 11(3): 173-182.
- [18] R. Din, A. Samsudin, T.Z.T. Muda, A. Amphawan. M.N. Omar. Fitness value based evaluation algorithm approach for text steganalysis model. *International journal of mathematic models and methods in applied science*. 2013; 7(5): 551-558.