Borneo International Journal eISSN 2636-9826; Vol. 6 (3); 2023; 96-103
Published by *Majmuah Enterprise*
www.majmuah.com

The Framework Design of Feature-based Method on Single-bit Technique Concealing Process using UML Representative

ROSHIDI DIN and SUNARIYA UTAMA
School of Computing, Universiti Utara Malaysia (UUM), 06010 Sintok, Kedah Darul Aman, MALAYSIA
Email: roshidi@uum.edu.my, sunariya.utama1@ahsgs.uum.edu.my| Tel: +60175981306, +601131548575 |

Received: September 14, 2023
Accepted: September 22, 2023
Online Published: September 26, 2023

Abstract

Text steganography is part of information hiding that focuses on concealing sensitive data within seemingly innocuous textual content. The feature-based method, which hides the hidden message based on the distinctiveness of each letter in the text, is taken into consideration in this research as a method categorization for text steganography. It integrates the three tecniques into a feature-based method that creates stego text utilising a single bit concealing mechanism in each letter. This paper introduces the three techniques single bit framework for feature-based method of text steganography within the Unified Model Language (UML). The UML that is represented diagram such as use case diagram, sequence diagram and class diagram. This UML representation of technique of feature-based method as the visualization in developing of text steganography in conceal the hidden message based on embedding and extracting implementation.

Keywords: Embedding process, Extracting process, Stego key

1. Introduction

The knowledge of art and science using hidden messages so that their existence cannot be seen by the human vision sense is known as steganography (Ahvanooey et al., 2019; Balu et al., 2019; Utama & Din, 2022) . If a person analyses an object that conceals cover information data using steganography, they won't suspect that it contains a hidden message and won't be enticed to decrypt the data. Secure private information is main key component of steganography when used as part of information concealment (Alshamsi et al., 2022; Saad et al., 2023). Since ancient times, steganography has been employed as a practical method. Greek terms steganos, which means covered or hidden, and graptos, which means to write or draw, are the origin of the term steganography. Steganography is the literal use of covered writing that dates back hundreds of years, long before the term "steganography" was coined. In short, Greek history, for instance, relates how Demeratus from Xerxes fashioned a hardwood basis for a wax tablet and later covered it with a fresh layer of wax in order to warn the Spartans of an impending invasion. Next, the other sample about steganography is when information was written on a piece of paper during World War II using invisible ink so that it wouldn't fall into the wrong hands. Simply put, it will appear to be a blank piece of paper (Selvigrija & Ramya, 2015; Shafi et al., 2017). The main benefit of steganography, based on the provided samples, it can be concluded that no one will be aware that there are archives with information buried within them, especially if the archives resemble other common files (Baawi et al., 2020; Ditta et al., 2020).

The steganography, which was divided into two categories. Technical steganography is the first method, which is utilised in non-media of text such as images, audio files, and other digitally indecipherable codes (Ciptaningtyas et al., 2019; Ditta et al., 2020; Taha et al., 2018). The second type of steganography uses natural language and implements the secret code through text (Xiang et al., 2020; Yang et al., 2019). In order to prevent a third party from learning about the existence of the concealed message in stego text, natural language steganography, the message is concealed in plain text. To put it another way, steganography in text can make the secret information difficult for others to notice or detect while still being conveyed to the correct recipients for them to understand (Al Azzawi, 2019; Saad et al., 2023). The use of natural language steganography hides messages in text. Text steganography and linguistic steganography are the two basic divisions. Text steganography refers to messages that manipulate text elements including words, spaces, lines, and other characters inside sentences(Torvi et al., 2016; Utama & Din, 2022). Meanwhile, linguistic steganography conceals messages by altering the data that was encoded using linguistic order (Nechta, 2018; Xiang et al., 2019). The Prisoner's Problem from Figure 1 could be used as an analogy to demonstrate how steganography works in the text domain.

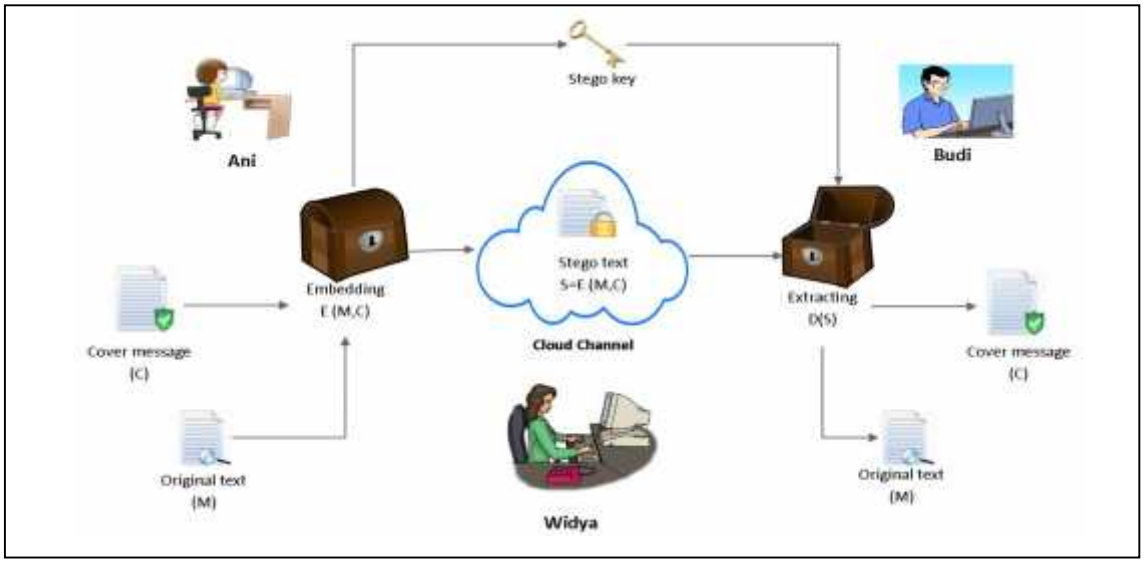

Figure 1: Analogy of text steganography processes

Figure 1 illustrates the similarity by Ani delivers an original text (M) and a cover message (C) in order to manage an embedding known as stego text (S) that contains a stego key (K). It is best to use the invertible function e: M, C, and S first. Ani can use a stego key (K) through e (M, C) = S to plan an original text (M). Widya won't be suspicious of S because it is a stego object and has an invertible function. In order to decode the procedure usage function and retrieve the original text M and cover message C using a stego key K, Budi will then establish that e - 1(S) = M, C. Widya won't be aware of the embedding and information extraction from stego text S procedure to ensure that steganography is successful (Mishra, 2015) .

Two approaches are used in the development of techniques: word-rule-based and feature-based methods (Arman et al., 2013; Din et al., 2018; Muhammad et al., 2021). By shifting words around in the text, the word-rule based technique uses word patterns to insert a hidden message. The techniques employed are word-shift coding and line-shift coding. Line-shift coding further obscures the concealed message by vertically shifting it between text lines. In contrast, word-shift coding conceals the concealed message by changing its length across words horizontally). Meanwhile, the method that altered a text's distinguishing feature characteristic based on a code word is referred to as feature-based. This approach uses a pattern letter or word length to cover a secret message while making it appear as though nothing has changed in the text (Dhawan & Gupta, 2021; Din et al., 2019; Li et al., 2017). The reader could be difficult in identifying the hidden information in this text due to the feature-based technique (Alanazi et al., 2020; Roy & Manasmita, 2011; Taka, 2021). As a result, numerous academics may employ feature-based methods based on the characteristics of languages used around the world (Din et al., 2019; Tu et al., 2018). This paper focus on the design of feature-based method that used in English letters in term of Unified Model Language (UML). It because UML diagram able to visualize the construction of programming language code, that become guidance the programmer to develop the system (Bunzel et al., 2021; Fauzan, 2019). It expected the technique scheme of feature-based method in form of UML promote the representing the implementation technique in order to better understand the main idea the information of development technique.

2. Methods

Implementation of feature-based method are able to implement in English text or A-Z letters (Wang et al., 2013). As part of the stego key, the concealed message is converted into binary during the feature-based method's development (Naharuddin et al., 2019). The stego key in a feature-based methodology embeds binary bits from a secret message into a text letter of choice. The main element of steganography that serves as a cover for the secret message is the stego key (Boroumand et al., 2019; Odeh et al., 2012). As a component of the feature-based methodology, this paperfocuses on three schemes that shared comparable characteristics for hiding the concealed message in the text. They are CURVE, VERT, and CALP (Dulera et al., 2011; Osman et al., 2014), which embed the concealed message by transforming it into a the text message can be hidden using a single binary bit (0 and 1). The cover text's letterswhich could include a hidden message make up the single binary bit of the approach used in this work. The feature-based text steganography algorithms are shown in Table 1.

Table 1: Scheme of Feature-based of one bit implementation

Group ID	Stego Key of Feature-based Technique		
ID (Bit)	**One binary bits**		
	CALP	**CURVE**	**VERT**
A (0)	"i, j"	"B, C, D, G, J,O, P, Q, R, S, U"	"A, C, G, H, M, N, O, Q, S, U, V"
B (1)	"A, a, c"	"A, E, F, H, I,K, L, M, N,T, V, W, X,Y, Z"	"B, D, E, F, I, J, K, L, P, R, T"

Based on Table 1, These three methods were picked because of their similar implementation strategies. The algorithms and group ID have been provided by previous researchers (Bhattacharyya et al., 2011; Dulera et al., 2011; Osman et al., 2014). The conversion procedure is based on the use of single binary bits in the schemes that three mentioned techniques were Vertical Straight Line (VERT), Curve in Character Subheading (CURVES), and Changing Alphabet Letter Patterns (CALP) (Dulera et al., 2011).

First, CALP mapping the binary attempts to change the pattern of many letters in the cover text throughout the embedding phase in order to manipulate the English letter sequence of the buried content. For the pattern, CALP is using an unutilized alphabetic symbol. This scheme's mapping sequences concentrate on the letters "i" and "j" because they contain a dot (.) that can be used to incorporate 0 bits. Following that, "a", "A", and "c" characters for embedding in 1 bits.

Second, the CURVE method categorises letters into two categories based on shape, or character is curved or not. The CURVE scheme, which serves as a stego key, is based on letters that are fully or partially curved; therefore, any word that contains a curve will be assigned the group ID (A), which designates that the letters "*B, C, D, G, J, O, P, Q, R, S, U*" conceal 0 bits of a hidden data. A letter that is completely straight is known as group ID (B), while letters like "*A, E, F, H, I, K, L, M, N, T, V, W, X, Y,* Z" conceal one bit of a concealed data.

Third, the VERT scheme separates each English letter into two groups while using a character's straight line as the foundation for each grouping. The VERT method separated the group letter into two group names similarly to CURVES. The letters "A, C, G, H, M, N, O, Q, S, U, V, W, Y, Z" and "B, D, E, F, I, J, K, L, P, R, T" are examples of Group ID (A) that do not have a vertical straight line and conceal 0 bits of secret data. Group ID (B) is a letter that contains a vertical line and conceals 1 bits of hidden data. As a result, these three methods can be used to change the grammar of both English and other languages that use the A-Z letters.

This paper focuses on the design of three feature-based methodology schemes for use case diagrams, sequence diagrams, and class diagrams in UML designs. These three diagrams of UML are expected to able to introduce the implementation of three techniques of feature-based method of single bit in concealing process of text steganography.

3. Design of Feature-based Method in UML Diagrams

The UML design serves as the basis for the development and system-level implementation of the methodology in the feature-based method (Fauzan, 2019). A system's process and the data flows into and out of the process are depicted graphically in the logical design.

3.1 Use Case Diagram

The simplest way to depict a user's engagement with a system is through a use case model, which demonstrates the relationships between Users and the different use cases that include them are discussed in Diagram & Aleryani's 2017 article. The demand in the user of the system's development using an existing technique that is feature-based is shown in Figure 2 as follow.

Borneo International Journal eISSN 2636-9826; Vol. 6 (3); 2023; 96-103
Published by *Majmuah Enterprise*
www.majmuah.com

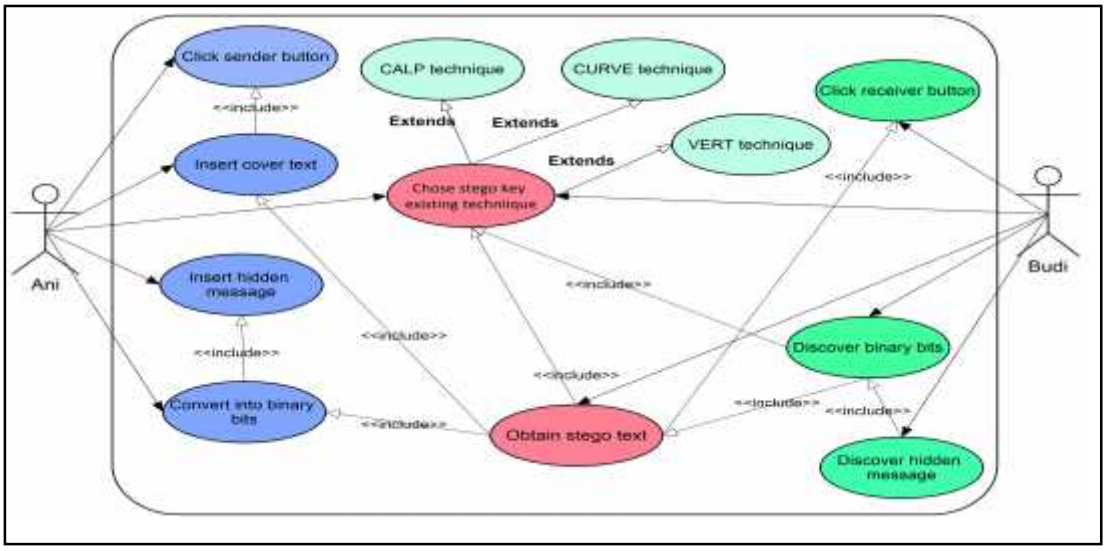

Figure 2: Use case diagram of existing techniques used

The prerequisites for the feature-based strategy used in this investigation are shown in Figure 2. In this use case diagram scenario, Ani, the sender, starts by clicking the sender button before inserting cover text and a secret message. The hidden message should then be translated to binary bits. The requirement is then to supply the three techniques strategies (CALP, CUVE, and VERT techniques) for choosing the stego key during the embedding process in order to retrieve the stego text. In the meantime, Budi as the receiver clicked the receiver button to find the stego text that had been retrieved using the method that Ani had previously chosen. After that, it converted the binary bit in order to re-discover the buried message's original text.

3.2 Sequence Diagram

Sequence diagram is an dealings diagram that demonstrates how the working steps work in harmony to depict how techniques flow through the system. It is a message sequence chart construct (Al-fedaghi, 2021). In Figure 3, it illustrates the sequence diagram of the existing of technique in embed the process of feature-based method of text steganography.

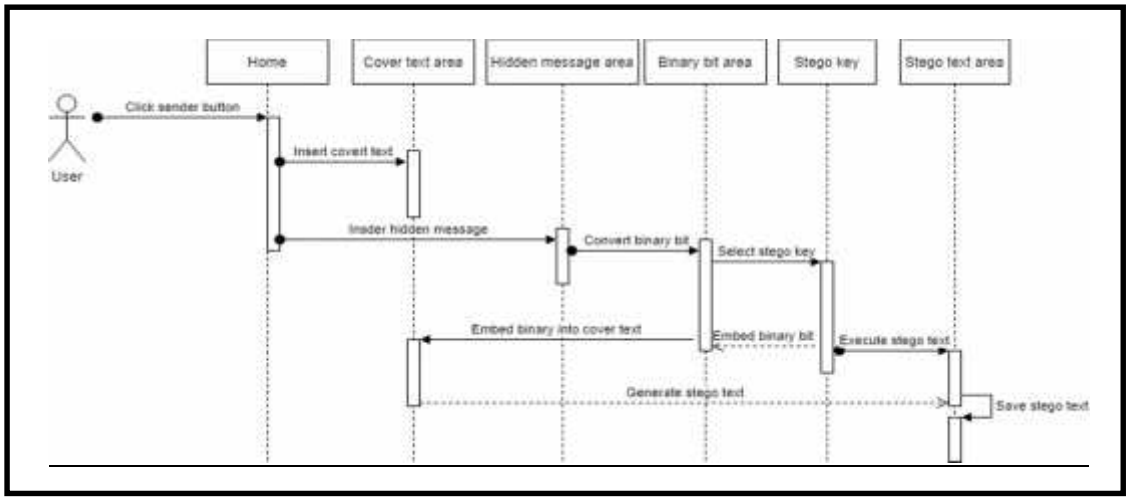

Figure 3: Sequence diagram of embedding process on existing techniques

Figure 3 shows the phase of embedding process in existing technique as the sender to achieve the stego text. In the sequence diagram, the sender is illustrated starting with their home screen, where they decide to place cover text and use that space as a text medium for a concealed message. After that, the hidden message is added and converted to a binary bit there. In addition, the actor selected one of the available feature-based methods to create the stego text that was placed in the area. This method embeds a binary bit in the cover text. The sequence then stores the stego text in plain text format. As a result, the feature-based method's embedding process is described in the sequence diagram. Next, the sequence diagram of extracting process of technique that is used receiver as the actor shows in Figure 4.

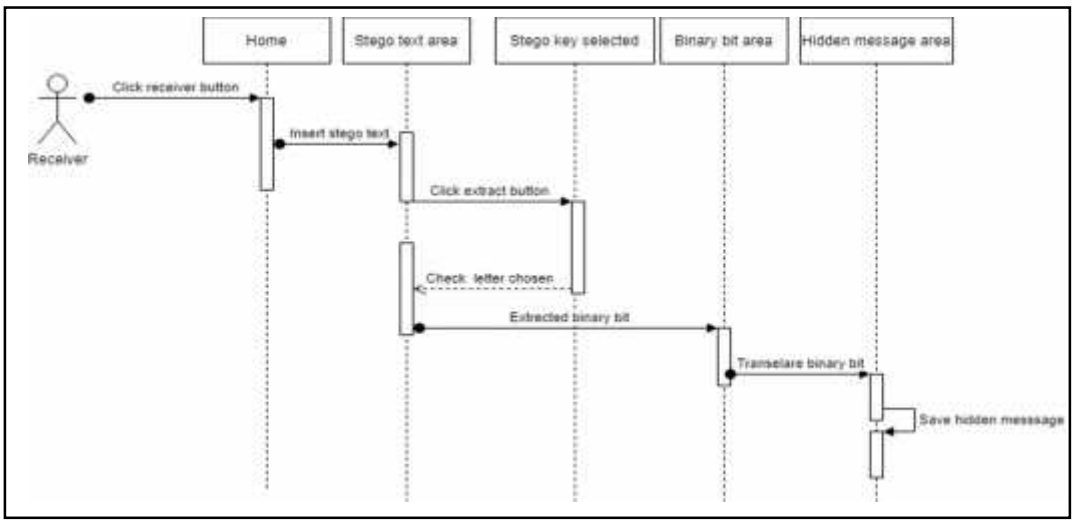

Figure 4: Sequence diagram of extracting process on existing techniques

Figure 4 depicts the stage of the extraction process using the sender technique to obtain the stego text. The process is shown in the sequence diagram as starting with the actor selecting the receiver in the home interface before inserting the stego text from the sender environment, which is saved in plain text format and shown in the stego text section. Then, it chooses the stego key that the sender employs in the sender environment to extract the binary bit from the stego text that is placed on the binary bit region. Once it has achieved binary bits, the receiver translates the hidden message that was displayed in the hidden message area.

3.3 Class Diagram

The structure of a system and the interactions between items are described in class diagrams, which are a sort of static structure diagram(Abdelnabi, 2020; Hoang et al., 2020). This diagram is able to assist conceptual detail model and structure of implantation technique as a guidance to develop the technique programming language. Figure 5 displays the class diagram of existing technique of feature-based method.

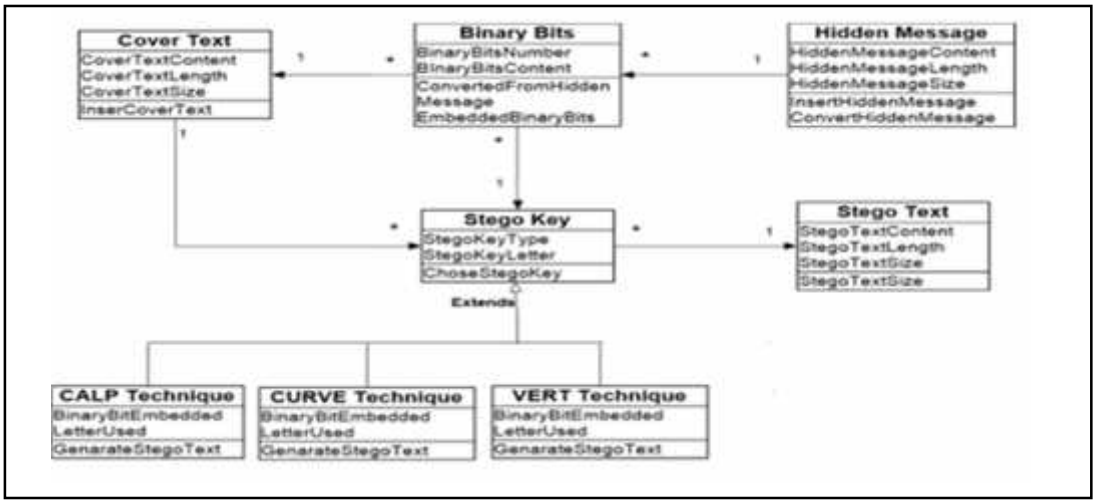

Figure 5: Class diagram of existing techniques

Figure 5 shows the class diagram that depicts the feature-based method's static implementation approach. The cover text class, secret message class, binary bit class, stego key class, and stego text class are the five main classes in this diagram. In terms of content, length, and size, the properties for cover text, hidden message, and stego text are similar. The binary bit can only have attributes of content and number, while the stego key can only have attributes of type and letter.

The class diagram is also an operation related to each class's relationship. The binary bits class and the hidden message have a relationship in which the latter indicates that the latter is translated into binary bits during feature-based implementation of the hidden message. Then, the stego key has a relationship with the binary bits and cover text, as well as an inheritance relationship with four existing methods: the CALP, CURVE, and VERT techniques classes. The letter utilised, action, and embedding of binary bits in stego language make up the attributes of the current method. The relationship between the stego key class and the stego text class includes both relationships and inheritance relationships with other classes. Those are the static view of the relationship class of the existing technique of feature-based method in the form of a class diagram.

4. Conclusions

In this paper, it integrates the three techniques on feature-based method of text steganography that utilize single bit based on embedding and extracting process in one letter in the text. The three technique that utilizes which are CALP, CURVE and VERT technique that divide the group single bit in form of 0 bit and 1 bit in converting the hidden message. Then, this paper presents the three techniques of feature-based method of text steganography using UML diagram representation. The use of UML diagrams was crucial during the design and development phases since it allowed for the clear separation of components and interactions as well as a visual representation of the system's architecture. The UML diagram includes a use case diagram to show the connections between users and use case requirements, a sequence diagram as an interaction diagram to show how a technique is used, and a class diagram to show the connections between classes, methods, and attributes. In addition to simplifying the implementation process, this UML diagram visual representation would be an invaluable resource for any future improvements or adjustments to the feature-based method of text steganography.

Acknowledgments

The authors would like to express our grateful to School of Computing of Universiti Utara Malaysia. This paper was supported by Universiti Utara Malaysia.

References

Abdelnabi, E. A. (2020). Generating UML Class Diagram using NLP Techniques and Heuristic Rules. *IEEE*. 277–282.

Ahvanooey, M. T., Li, Q., Hou, J., Rajput, A. R., & Chen, Y. (2019). Modern text hiding, text steganalysis, and applications: A comparative analysis. *Entropy*, *21*(4), 1–29. https://doi.org/10.3390/e21040355

Al-fedaghi, S. (2021). UML Sequence Diagram : An Alternative Model. *ArXic preprint*, *12*(5), 635–645.

Al Azzawi, A. F. (2019). A Multi-Layer Arabic Text Steganographic Method Based on Letter Shaping. *International Journal of Network Security & Its Applications*, *11*(01), 27–40. https://doi.org/10.5121/ijnsa.2019.11103

Alanazi, N., Khan, E., & Gutub, A. (2020). Inclusion of Unicode Standard seamless characters to expand Arabic text steganography for secure individual uses. *Journal of King Saud University - Computer and Information Sciences*, *xxxx*. https://doi.org/10.1016/j.jksuci.2020.04.011

Alshamsi, A., Albaloushi, S., Alkhoori, M., Almheiri, H., & Ababneh, N. (2022). Enhancing Arabic Text Steganography Based on Unicode Features. *International Journal of Computing and Digital Systems*, *11*(1), 685–693. https://doi.org/10.12785/ijcds/110155

Altigani, A., & Barry, B. (2013). A hybrid approach to secure transmitted messages using advanced encryption standard (AES) and Word Shift Coding Protocol. *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference On*, 134–139.

Arman, A. A., N, A. B. P., Purwarianti, A., & Kuspriyanto. (2013). Syntactic Phrase Chunking for Indonesian Language. *Procedia Technology*, *11*(November 2015), 635–640. https://doi.org/10.1016/j.protcy.2013.12.239

Baawi, S. S., Nasrawi, D. A., & Abdulameer, L. T. (2020). Improvement of "text steganography based on unicode of characters in multilingual" by custom font with special properties. *IOP Conference Series: Materials Science and Engineering*, *870*(1). https://doi.org/10.1088/1757-899X/870/1/012125

Balu, S., Babu, C. N. K., & Amudha, K. (2019). Secure and efficient data transmission by video steganography in medical imaging system. *Cluster Computing*, *22*, 4057–4063. https://doi.org/10.1007/s10586-018-2639-4

Bhattacharyya, S., Indu, P., Dutta, S., Biswas, A., & Sanyal, G. (2011). Hiding Data in Text Through Changing in

Alphabet Letter Patterns (CALP). *Journal of Global Research in Computer Science*, *2*(3).

Boroumand, M., Chen, M., & Fridrich, J. (2019). Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, *14*(5), 1181–1193. https://doi.org/10.1109/TIFS.2018.2871749

Bunzel, N., Steinebach, M., & Liu, H. (2021). Cover-aware Steganalysis. *Journal of Cyber Security and Mobility*, *10*, 1–26. https://doi.org/10.13052/jcsm2245-1439.1011

Ciptaningtyas, H. T., Anggoro, R., & Aji Fadhillah, M. B. (2019). Text steganography on sundanese script using improved line shift coding. *International Electronics Symposium on Knowledge Creation and Intelligent Computing, IES-KCIC 2018 - Proceedings*, 254–261. https://doi.org/10.1109/KCIC.2018.8628471

Dhawan, S., & Gupta, R. (2021). Analysis of various data security techniques of steganography: A survey. *Information Security Journal*, *30*(2), 63–87. https://doi.org/10.1080/19393555.2020.1801911

Aleryani, A. Y. (2017). Comparative Study between Data Flow Diagram and Use Case Diagram. *International Journal of Scientific and Research Publication*. *February*.

Din, R., Bakar, R., Utama, S., Jasmis, J., & Elias, S. J. (2019). The evaluation performance of letter-based technique on text steganography system. *Bulletin of Electrical Engineering and Informatics*, *8*(1), 291–297. https://doi.org/10.11591/eei.v8i1.1440

Din, R., Utama, S., & Mustapha, A. (2018). Evaluation review on effectiveness and security performances of text steganography technique. *Indonesian Journal of Electrical Engineering and Computer Science*, *11*(2), 747–754. https://doi.org/10.11591/ijeecs.v11.i2.pp747-754

Ditta, A., Azeem, M., Naseem, S., Gulzar Rana, K., Adnan Khan, M., & Iqbal, Z. (2020). A secure and size efficient algorithm to enhance data hiding capacity and security of cover text by using unicode. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2020.07.010

Dulera, S., Jinwala, D., & Dasgupta, A. (2011). Experimenting with the Novel Approaches in Text Steganography. *International Journal of Network Security & Its Applications*, *3*(6), 213–225. https://doi.org/10.5121/ijnsa.2011.3616

Fauzan, R. (2019). Use Case Diagram Similarity Measurement : A New. *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 3–7.

Hoang, M., Vo, L., & Hoang, Q. (2020). *Transformation of UML class diagram into OWL Ontology Transformation of UML class diagram into OWL Ontology*. *1839*. https://doi.org/10.1080/24751839.2019.1686681

Li, S., Jia, Y., & Kuo, C. C. J. (2017). Steganalysis of QIM Steganography in Low-Bit-Rate Speech Signals. *IEEE/ACM Transactions on Audio Speech and Language Processing*, *25*(5), 1011–1022. https://doi.org/10.1109/TASLP.2017.2676356

Mishra, R. (2015). A Review on Steganography and Cryptography. *Computational Intelegence in Machine Learning*, 119–122.

Muhammad, M. H., Hussain, H. S., Din, R., Samad, H., & Utama, S. (2021). *Review on feature-based method performance in text steganography*. *10*(1), 427–433. https://doi.org/10.11591/eei.v10i1.2508

Naharuddin, A., Wibawa, A. D., & Sumpeno, S. (2019). A High Capacity and Imperceptible Text Steganography Using Binary Digit Mapping on ASCII Characters. *Proceeding - 2018 International Seminar on Intelligent Technology and Its Application, ISITIA 2018*, 287–292. https://doi.org/10.1109/ISITIA.2018.8711087

Nechta, I. V. (2018). New Steganalysis Method for Text Data Produced by Synonym Run-Length Encoding. *2018 14th International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering, APEIE 2018 - Proceedings*, 188–190. https://doi.org/10.1109/APEIE.2018.8545230

Odeh, A., Alzubi, A., Hani, Q. B., & Elleithy, K. (2012). Steganography by multipoint Arabic letters. *2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, 1–7. https://doi.org/10.1109/LISAT.2012.6223209

Osman, B., Din, R., & Idrus, M. R. (2014). Capacity Performance Of Steganography Method In Text Based Domain. *X*(X), 1–7.

Roy, S., & Manasmita, M. (2011). A novel approach to format based text steganography. *Proceedings of the 2011 International Conference on Communication, Computing & Security - ICCCS '11*, 511. https://doi.org/10.1145/1947940.1948046

Saad, E., Al, N., & Algamdi, A. (2023). *S*urvey steganography applications. *Innovative Research Publication*, 69–75.

Selvigrija, P., & Ramya, E. (2015). Dual steganography for hiding text in Video by Linked List Method. *2015 IEEE International Conference on Engineering and Technology*, 1-5.

Shafi, I., Noman, M., Gohar, M., Ahmad, A., Hassan, S., Jamil, A., Khan, M., & Din, S. (2017). An adaptive hybrid fuzzy-wavelet approach for image steganography using bit reduction and pixel adjustment. *Soft Computing*. https://doi.org/10.1007/s00500-017-2944-5

Taha, A., Hammad, A. S., & Selim, M. M. (2018). A high capacity algorithm for information hiding in Arabic text. *Journal of King Saud University - Computer and Information Sciences*, *32*(6), 635-754.

Taka, F. R. S. (2021). Text Steganography based on Noorani and Darkness. *Journal Information Hiding Multimedia*, *12*(3), 127–139.

Torvi, S. D., Kumar, K. B. S., & Das, R. (2016). An Unique Data Security using Text Steganography. *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on IEEE*, *2016.*, 3834–3838.

Tu, S., Huang, X., Huang, Y., Waqas, M., & Rehman, S. U. (2018). SSLSS: Semi-supervised learning-based steganalysis scheme for instant voice communication network. *IEEE Access*, *6*, 66153–66164. https://doi.org/10.1109/ACCESS.2018.2879328

Utama, S., & Din, R. (2022). Performance Review of Feature-Based Implementation Text Steganography Approach Method. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *2*(2), *2*(2), 325–333.

Wang, F., Huang, L., Chen, Z., Yang, W., & Miao, H. (2013). A Novel Text Steganography by Context-Based Equivalent Substitution. *Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference*, 1–6.

Xiang, L., Guo, G., Yu, J., Sheng, V. S., & Yang, P. (2019). A convolutional neural network-based linguistic steganalysis for synonym substitution steganography. *Mathematical Biosciences and Engineering*, *17*(January), 1041–1058.

Xiang, L., Yang, S., Liu, Y., Li, Q., & Zhu, C. (2020). Novel linguistic steganography based on character-level text generation. *Mathematics*, *8*(9), 1–18. https://doi.org/10.3390/math8091558

Yang, Z., Huang, Y., & Zhang, Y. J. (2019). A Fast and Efficient Text Steganalysis Method. *IEEE Signal Processing Letters*, *26*(4), 627–631. https://doi.org/10.1109/LSP.2019.2902095