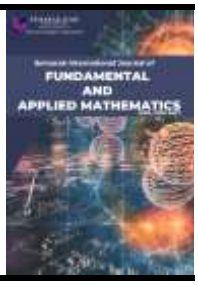
Semarak International Journal of
Fundamental and Applied Mathematics

Journal homepage:
https://semarakilmu.my/index.php/sijfam
ISSN: 3030-5527



A Mathematical Formulation of Classified Variables for Enhanced Detection in Natural Language Steganalysis

Siti Norussaadah Mohd Salleh[1*], Roshidi Din[1]

[1] School of Computing, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

ARTICLE INFO

Article history:
Received 13 January 2024
Received in revised form 15 February 2024
Accepted 18 March 2024
Available online 30 March 2024

Keywords:
Natural language steganalysis; classified variables; hidden information; formulated key; text conversation

ABSTRACT

Steganalysis, which is the counterpart to steganography, is an art and science that is devoted to detecting hidden information that has been concealed within seemingly harmless digital media. With the advancement in technology nowadays, the techniques used in steganography also become more complex, making it necessary to continuously improve steganalysis methods to keep up with emerging threats and ensure digital data's confidentiality, authenticity, and privacy. Steganalysts face the challenge of uncovering hidden information within covert media, requiring analysis of both original and altered media. While efficient steganalysis tools exist for specific approaches, devising a universal solution for all steganography techniques remains challenging. With the utilization of the formulated key, this paper aims to classify and analyze variables used in the steganalytic system. Thus, three views of classified variables are presented to address the pattern of detection. These are known as trade-off value-based, probability of character variable, and Support Vector Machine-based (SVM-based). Hence, it is expected that this scheme will become one of the alternative ways to enhance the steganalytic system for discovering the hidden message in communication. Furthermore, it is suggested the necessity for enhanced steganalysis techniques and tools, as well as the significance of academic and professional education in this area.

1. Introduction

One of the concern areas with information hiding is steganalogy which has attracted more attention since last decade [1,2]. It has played an important role in secret communication and information [3,4] such as medical and healthcare security [5], e-military [6], and also forensics [7] matters from medieval times through the 20th century. In fact, steganalogy is the technique for digitally covering and detecting information through a secret communication channel. Steganalogy guarantees the cover messages are perceptually unchanged after concealing and detecting the covered writing. Steganalogy can be classified into two parts which are steganography and steganalysis.

[*] *Corresponding author.*
E-mail address: ssaadah84@gmail.com

https://doi.org/10.37934/sijfam.1.1.4961b

Steganography uses numerous carriers for transferring information. For instance, digital steganography consists of carriers such as image steganography [8], audio [9], and video [10] which have produced good results. Steganography plays an important role in preserving the confidentiality and integrity of the data. It aims to prevent the detection of secret data when exchanging messages in communication [11]. There are several techniques and methods that can be employed in steganography using steganographic tools. For example, OpenStego [12], SSuite Picsel [13], StegProxy [14], and Steghide [15] have been developed based on digital steganography.

Meanwhile, steganalysis is reflected as the discovery of secret information [16] where the main goal is to reveal the hidden information that exists in the particular data. It can be categorized into digital steganalysis and natural language steganalysis. Natural language steganalysis can be classified further into text-based and linguistic-based. In contrast, digital steganalysis comprises of image, audio, video, and network. There have been numerous studies done on steganalysis in the digital domain based on past research. For example, there are studies on image steganalysis [17-19] , audio steganalysis [20,21] and video steganalysis [22,23]. The Simmons idea based on the "Prisoners' problem [24] is shown in Figure 1 below.

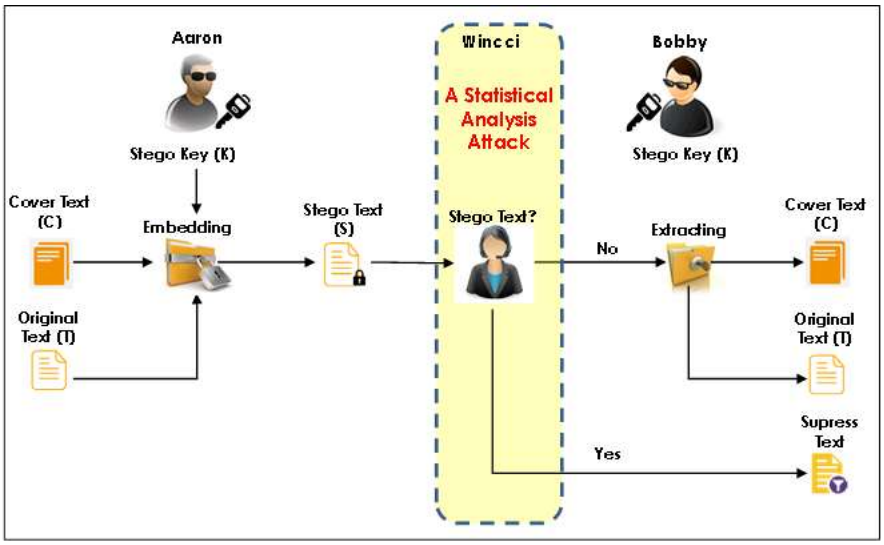


Fig. 1. A visualize of text steganalytic model

1.1. Text Steganalysis

Nowadays, with the growing interest in online communication, text steganalysis is needed to detect the existence of hidden message in the features characteristics, statistical probabilities, or linguistic structures of the online text conversation. In statistical-based, text steganalysis will attempt to locate the reliable anticipated secret text in that particular natural language text domain [25]. The format-based attack is known as the steganalysis algorithm that is used on text steganalysis. It is shown that the algorithm mostly will alter the statistical features of carriers. There are several types of techniques that can be associated with this attack. For instance, the technique that is based on tags that have been proposed [26] is used to discover the hidden message embedded in tags of a webpage. The basic idea of this technique is to check for a mismatch that occurs in the tags. Another technique that has been studied [27] is character substitution. Basically, this technique examines the language characters and language commas in order to identify the characteristic errors in the texts that contain two languages. Therefore, the main idea of this paper is to address a formulation of a

technique in the natural language steganalysis domain. Ordinarily, the strength of the selected steganalytic framework model has an impact on the techniques in steganalysis [28].

2. Classified Variables in Formulated Steganalytic Key

The fact of the formulated key condition is depending on variables used during the embedding process of any steganalytic system. Indeed, the strength of variables has an impact on the formulated key used for the steganalysis method. Actually, there are three types of classified variables have been identified which are trade-off value-based, probability of character variable, and Support Vector Machine-based (SVM-based).

2.1 Trade-off Value-based

Based on the idea of the "Prisoners' problem", it is assumed that online communication has occurred between two criminals named Aaron and Bobby who had been detained in Azkaban prison. They have been isolated between two fortresses namely the Northern cell and the Southern cell. They are only allowed to communicate online via Wincci, as an online System Administrator. Therefore, Wincci as a caretaker tries to monitor the communication of these two criminals in order to discover whether there is any hidden message in their online communication. One of the significant variables that can be used by Wincci is based on tag character. It is done by performing tag-mismatch analysis on the online text conversation. There are two types of tag-mismatch that need to be considered which are tag-pair mismatch s_1 and null-tag mismatch s_2 [26]. Therefore, Wincci will try to analyze any information from the text conversation to identify any tag-mismatch of file size, *filesize* of a text conversation, and average of tags character, *k* in online communication so that she is able to calculate the embedded rate, *X* as following Eq. (1);

$$X = \frac{k \ x \ (2 \ x \ s_1 + s_2)}{filesize} \tag{1}$$

The value of embedded rate, *X* is to indicate the ratio of the length of the secret message to the file size. This value is calculated and compared with the decision threshold, α [26] by Wincci to determine the communication status. Actually, she believes that if the embedded rate, *X* is greater than the decision threshold, α, the online text conversation can be identified as stego communication where this communication is embedded with a secret message and otherwise is recognized as normal communication. Therefore, the status of communication can be interpreted as Eq. (2);

$$\begin{array}{ll} X \geq \alpha & \text{stego communicat ion} \\ X < \alpha & \text{normal communicat ion} \end{array} \tag{2}$$

Besides that, Wincci also can discover the hidden message of text conversation in online communication through the pattern of text communication. Through this idea, Wincci will try to calculate the randomness value, *H(x)* [29] of text communication patterns based on the binary code string of that text.

On the other hand, Wincci can manipulate the feature of the inherent characteristic of tag offset, *offset* [30] of a text conversation. Then, the value of tag offset, $offset_{value}$ will be determined through the alteration function of each character in a text conversation between Aaron and Bobby. This tag offset value, $offset_{value}$ can be represented by two values either nonpositive or nonnegative. If Wincci

finds that the feature of the inherent characteristic of text conversation decreases, $offset_{decrease}$, this communication is identified as stego communication. Otherwise, if it increases, $offset_{increase}$, the communication is normal. This situation can be formulated as following Eq. (3);

$$\begin{array}{ll} offset_{decrease} & \text{stego communication} \\ offset_{increase} & \text{normal communication} \end{array} \tag{3}$$

2.2 Probability of Character Value-based

Another view of natural language steganalysis perspective that can be illustrated is through the probability of character variables. Through this view, Wincci can use the feature of the probability of position character of a text conversation to discover any secret message in the online communication. It can be done by calculating the similarity coefficient, α [31] as follows Eq. (4);

$$\alpha = 1 - \frac{\sum_{i=1}^{26} |\rho_1(i) - \rho_2(i)|}{\sum_{i=1}^{26} |\rho_1(i) + \rho_2(i)|} = 1 - \frac{1}{2}\sum_{i=1}^{26} |\rho_1(i) - \rho_2(i)| \tag{4}$$

where,

i	1, 2, … 26 (represents the 26 letters of alphabetic characters *A* to *Z*)
$\rho 1(i)$	probability of each letter involve in a group of alphabetic characters *A* to *Z*
$\rho 2(i)$	probability of each text conversation involves in group of alphabetic characters *A* to *Z*

After that, Wincci will compare the value of the similarity coefficient, α with the decision threshold, Δ [31]. If the value of the similarity coefficient, α is greater than the value of the decision threshold, Δ of the online text conversation, this communication is classified as stego communication. In different circumstances, the value of similarity coefficient, α is lower than the decision threshold, Δ value, the communication is identified as normal communication. Therefore, this situation can be formulated as follows Eq. (5);

$$\begin{array}{ll} \alpha \geq \Delta & \textit{stego communication} \\ \alpha < \Delta & \textit{normal communication} \end{array} \tag{5}$$

In addition, Wincci can use the feature of probability of space characters to determine the existence of the hidden message in text conversation on online communication. Wincci will try to calculate the probabilities of the space characters, P_1 and the probabilities of continuous space characters, P_2 [32] using Eq. (6) and Eq. (7) as below;

$$P_1 = \frac{number\,of\,all\,characters}{number\,of\,space\,characters} \tag{6}$$

$$P_2 = \frac{number\,of\,all\,space\,characters}{number\,of\,all\,continuous\,space\,characters} \tag{7}$$

Then, Wincci does the comparison between the values of both probabilities with the decision threshold, Δ in order to identify that communication. She believes that if P_1 and P_2 are greater than the decision threshold, Δ, the analyzed text conversation is considered as stego communication and otherwise as stated in Eq. (8) below;

$$\begin{array}{ll} P_1, P_2 \geq \Delta & \text{stego communicat ion} \\ P_1, P_2 < \Delta & \text{normal communicat ion} \end{array} \tag{8}$$

2.3 SVM-based

One of the most used views on natural language steganalysis perspective is represented through the SVM-based view. One of the views in SVM-based that can be used by Wincci is through character-based manipulation. The first character-based manipulation is the frequency of the character. Let's say SVM is used in order to detect suspicious text conversations between Aaron and Bobby in online communication. Context maximum rate, λ, and context maximum deviation, ϑ [33] can be used by SVM as a *trade-off* due to classifying between normal communication and stego communication. Therefore, this situation can be formulated as Eq. (9);

$$\begin{aligned} \lambda &= \frac{1}{n}\sum_{i=0}^{n-1}[\gamma_i = \gamma_{i,\max}] \\ \theta &= \frac{1}{n}\sum_{i=0}^{n-1}(\gamma_i = \gamma_{i,\max})^2 \end{aligned} \tag{9}$$

where,

i number of fitness
n number of context fitness
γ_i context fitness, either 0 or 1
$\gamma_{i,max}$ maximum context fitness, either 0 or 1

Furthermore, Wincci can utilize character-based manipulation through the frequency of text characteristics, f_k of text conversation, t_c between Aaron and Bobby. She will try to discover the hidden message based on that conversation. Through SVM, the Natural Relative Frequency (NRF) score [34] will be used to justify the status of communication either normal communication or stego communication. The communication status will be obtained by two parameters of NRF score which are known as expected value, α of NRF score, and variance value, γ of NRF score. Therefore, the value of NRF score, NRF_k^{tc}, expected value, α of NRF score, and variance value, γ of NRF score can be calculated through Eq. (10) and Eq. (11);

$$NRF_k^{tc} = K * \left(\frac{f_{k,i}}{\sum_{j=0}^{m_{k-1}} f_k} \right), 0 \leq i < m_k \tag{10}$$

$$\alpha = \frac{1}{n}\sum_{i=0}^{n-1} c_i$$
$$\gamma = \frac{1}{n}\sum_{i=0}^{n-1}\left(\frac{c_i - \alpha}{\alpha}\right)^2 \quad (11)$$

where,

i	number of scores
n	number of NRF scores
K	number of text characters
f_k	frequency of no of text character
m_k	maximum list of text character
c_i	occurring NRF scores

Besides, another character-based manipulation that can be used by Wincci is through character substitution. Ideally, Wincci will try to justify the ratio of abnormal character to normal character, (RAN) [27]. This RAN ratio will be used by SVM as a *trade-off* between normal communication and stego communication.

The next character-based manipulation that can be used by Wincci to discover the secret message of a text conversation in online communication is based on perplexity. Through this idea, Wincci will try to calculate the perplexity value, *P(w)* [35] based on the number of values, w_i, and the number of n-grams in the text, *N* which then passed to SVM for classification between normal communication and stego communication. The value of perplexity can be formulated as follows Eq. (12);

$$P(w) = \sqrt[N]{\prod_{i=1}^{N}\frac{1}{P(w_i \mid w_{i-2}, w_{i-1})}} \quad (12)$$

In addition, Wincci also can use context information in the text conversation in order to detect the presence of hidden information through the Vote Percentage, and *VP* calculation [36]. Then, SVM will use the *VP* value to classify whether the communication between Aaron and Bobby is normal communication or stego communication.

Moreover, Wincci is able to discover the hidden message of a text conversation in online communication through font attributes of text characters. Based on the distance of font attributes between every two adjacent characters, D_i, and Wincci will try to calculate the attribute distance frequency, S_k [37] of the text character. The value of the distance of font attributes between every two adjacent characters, D_i, can be obtained through Eq. (13) as below.

$$D_i = |P(t_i) - P(t_j)|$$
$$= |p(t_i) - p(t_j) + \gamma(t_i) - \gamma(t_j) + \delta(t_i) - \delta(t_j)| \quad (13)$$

where,

i	number of frequency
j	i + adjacent character ($i+1$)
n	number of distance frequency
$p(t_i)$	normal font attribute value
$p(t_j)$	normal font attribute value (with adjacent character)

$\gamma(t_i)$ variation of font attribute value
$\gamma(t_j)$ variation of the font attribute value (with the adjacent character)
$\delta(t_i)$ noise signal
$\delta(t_j)$ noise signal (with adjacent character)

Then, she will use this attribute distance frequency, S_k as a *trade-off* which can be manipulated by SVM in order to justify the communication status of either normal communication or stego communication. Therefore, this frequency can be formulated as shown in Eq. (14);

$$S_k = \sum_{i=0}^{n-2} D_i = k \tag{14}$$

Besides the character-based manipulation point of view, Wincci can use word-based manipulation to detect any hidden information. One of the word-based manipulations is based on word distribution where Spread Degree, *SD* [38] is computed as an unbalanced word location measurement. Then, two parameters which are the average of Spread Degree $\overline{SD}$, and variance of Spread Degree $Var(SD)$, are used by SVM to justify the text conversation as a normal or stego communication. This situation can be formulated as following Eq. (15);

$$\begin{aligned} \overline{\mathrm{SD}} &= \sum_{\mathrm{i}=0}^{\mathrm{m}} \mathrm{SD}(\mathrm{w}_{\mathrm{i}}^{'}) \frac{\mathrm{n_i}}{\mathrm{n}} \\ \mathrm{Var(SD)} &= \sum_{\mathrm{i}=0}^{\mathrm{m}} \left(\mathrm{SD}(\mathrm{w}_{\mathrm{i}}^{'}) - \overline{\mathrm{SD}} \right)^2 \frac{\mathrm{n_i}}{\mathrm{n}} \end{aligned} \tag{15}$$

where,
i number of words
m maximum of word
$w_i^{'}$ $0 \le i \le m$, is the (i+1)th word of the text
n_i number of occurrences

The second word-based manipulation is word correlation. Through this idea, Wincci will identify the N-Window Variance of Mutual Information (N-WVMI) matrix, V value, and Partial Average Distance, $D_{\alpha,K}$ value [39] to be used by SVM as a *trade-off* in order to differentiate between normal communication and stego communication. Actually, this idea can be stated through Eq. (16);

$$\begin{aligned} V &= \frac{1}{I} \sum_{i=0}^{M} \sum_{j=0}^{M} (S_{ij} - T_{ij})^2 \delta(i, j) \\ D_{\alpha,K} &= \frac{1}{K} \sum_{i=0}^{M} \sum_{j=0}^{M} |S_{ij} - T_{ij}| [|S_{ij} - T_{ij}| > \alpha] \lambda_k(i, j) \end{aligned} \tag{16}$$

where,
i number of items
j number of pair
M count of words in the word dictionary

I	pairs of items
α	a threshold of the distance of two N-WMI values
K	calculation of the first greatest item of matrix S_{MXM}
Sij	the pair of text segments in position (i,j)
Tij	the pair of training corpus in position (i,j)
$[\mid S_{ij} - T_{ij} \mid > \alpha]$	will be equal to 1, otherwise 0, and
$\lambda_k(i, j)$	will be equal to 1 if S_{ij} is the greatest K, otherwise 0

Meanwhile, the third word-based manipulation is the feature of word entropy of text conversation. A Detection Information, DI_i [40] is attained by applying Eq. (17). After that, Wincci will use these values in order to obtain the detection entropy, DE, and variance of detection entropy, Var(DE) as following Eq. (18);

$$S_i = \frac{1}{C}\left(\sum_{i=1}^{n} i\right)$$
$$DI_i = \log \frac{1}{S_i} \tag{17}$$

$$DE = \sum_{i=0}^{N-1} S_i \, DI_i = -\sum_{i=0}^{N-1} S_i \log S_i$$
$$Var(DE) = \sum_{i=0}^{M} S_i \left(DI_i - DE\right)^2 \tag{18}$$

where,

i	number of occurrences
n	number of words
S_i	score for word i
C	total of occurrences of all words

Then, Wincci will use these two values through SVM in order to classify whether it is a normal or stego communication.

Fourthly, Wincci also can use word-based manipulation that emphasizes the distribution of word frequency where the distribution of words will determine the accuracy of structure style. She will try to analyze two parameters of natural frequency zoned (NFZ) which are the average of NFZ, α_k value, and variance of NFZ, γ_k value [41]. These two parameter values can be utilized by SVM in order to detect the text conversation whether contains or does not contain a hidden message. These two values can be formulated as following Eq. (19);

$$\alpha_k = \frac{1}{n_k + 1} \sum_{i=0}^{n_k} d_{i,i-1}^{(k)} = \frac{1}{n_k + 1}(1 - 0) = \frac{1}{n_k + 1}$$
$$\gamma_k = \frac{1}{n_k + 1} \sum_{i=0}^{n_k} (d_{i,i-1}^{(k)} - \alpha_k)^2 \tag{19}$$

where,

i number of frequency
n_k number of NFZ frequency
$d_{i,i-1}^{(k)}$ distance of the (i+1)th and ith words in NFZ

Finally, word-based manipulation that can be used by Wincci is based on word shift which uses the neighbor difference concept [42]. Similarly, SVM can use this word-based manipulation to justify the text conversation as either normal communication or stego communication.

Another point of view that can be used by Wincci is language-based manipulation. One of these manipulations is the feature of Language Modelling (LM) of text conversation [43]. There are five parameters that can be utilized by Wincci which are word statistics, minimum n-gram context length, statistics of model probability estimations for n-grams in the sentence, model statistics for log probability of n-grams in the sentence, and total probability of the sentence. These parameters can be integrated as a single vector. Therefore, Wincci can use this single vector through the SVM method in order to justify the text conversation of online communication between Aaron and Bobby.

Additionally, Wincci can utilize the segment of sentences of a text conversation. There are two values that can be obtained through a segment of the sentence which are the Degree of Machine Reversibility (DMR) and Degree of Machine Preference (DMP) [44]. Therefore, Wincci will use these two values as a *trade-off* to the SVM method in order to detect the existence of the hidden message in the text conversation between Aaron and Bobby.

It can be concluded that Wincci as a steganalyst can utilize all the classified variables consisting of trade-off value-based, probability of character variable, and SVM-based as the views of natural language steganalysis perspective. These classified variables can be used to discover whether there is any hidden message in any communication that occurs between the cyber communities in this electronic era.

3. Challenges and Future Works

There is considerable research conducted on steganography by scholars, cyber experts, and digital forensic examiners. However, this is against the field of steganalysis as it poses a very challenging task for them due to the lack of observed output for steganalysis techniques and tools [45]. As digital communication transforms daily life, the widespread availability of steganography tools online enables individuals with minimal or no technical expertise to engage in hiding the information. These readily accessible tools are so user-friendly that anyone can conceal a secret message within an object with a single click, requiring no processing time [6]. Since shared data are susceptible to data leaks [46], the requirement for additional security measures to avoid information being obtained from questionable sources [47]. Moreover, the use of steganography for illegal purposes could pose a serious threat to national security. Thus, the main challenge faced by the steganalyst is to find hidden information in covert media and extract it as understanding the original media is essential to the steganalysis process. It involves analyzing the features and characteristics of both cover and stego media. According to research by [48], several obstacles confront steganalysis techniques such as known only the stego media, known only the hidden media in scattered form, and known the cover and stego media. To counter this threat, there is a need for improved steganalysis techniques and tools, as well as the implementation of a steganalysis system capable of effectively inspecting suspicious information. Although it is possible to develop efficient steganalysis tools for certain approaches, it is difficult to come up with a scheme that works for all steganography techniques. In the future, it is suggested to increase the volume of academic publications in this

domain to underscore its significance and widespread interest. It is crucial to promote professional and academic education in information hiding, enhancing comprehension, and refining methodologies within the field. In addition, future work can be extended to improve steganalysis methods or instruments that can eventually handle different steganography classes.

4. Conclusion

This study has addressed a formulation of a technique in the natural language steganalysis domain. Based on the mathematical formulation, all the variables are analyzed and classified according to the pattern of detection. The detection pattern is considered in order to distinguish all the variables utilized in the process of discovering the hidden message in the online text conversation. Hence, three categories of classified variables have been identified which are trade-off value-based, probability of character variable, and SVM-based as shown in Figure 2.

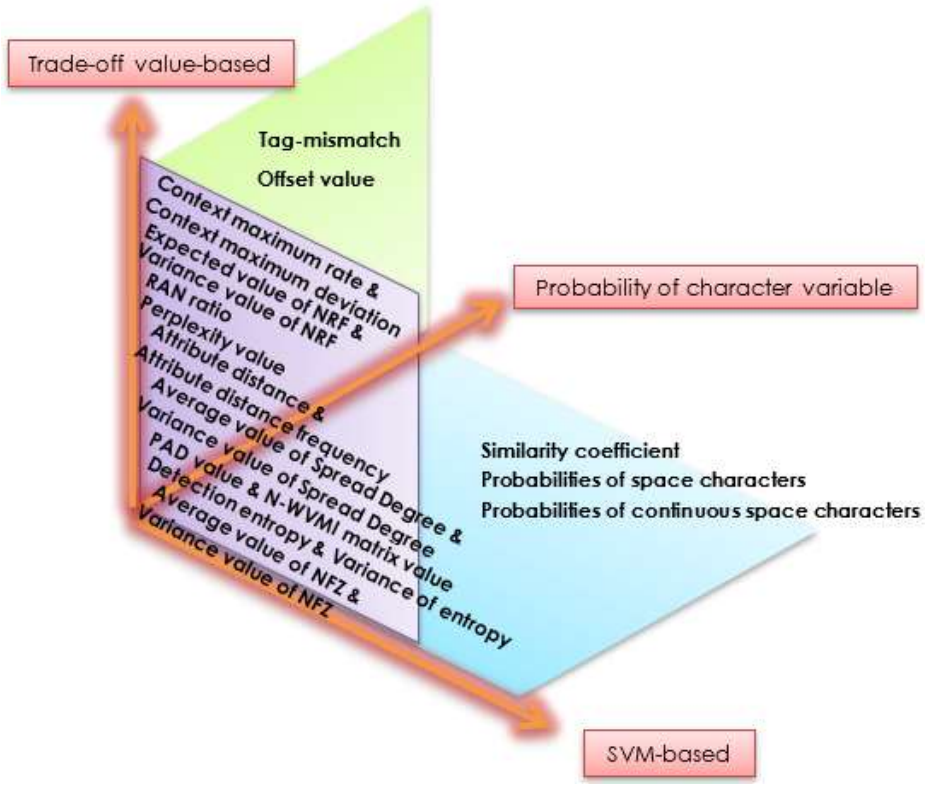


Fig. 2. Three views of classified variables

Through this view, it is assumed that a new scheme based on numerical representation will be proposed in the future effort. Thus, it is expected that this scheme will become one of the alternative ways to detect secret information in communication.

Acknowledgment

This research was not funded by any grant.

References

[1] Yadav, Pooja, and Sangeeta Dhall. "Comparative analysis of steganography technique for information security." *International Journal of Mathematical Sciences and Computing* 6, no. 4 (2020): 42–69. https://doi.org/10.5815/ijmsc.2020.04.05

[2] Shehab, Doaa A., and Mohmmed J. Alhaddad. "Comprehensive survey of multimedia steganalysis: Techniques, evaluations, and trends in future research." *Symmetry* 14, no. 117 (2022): 1–26. https://doi.org/10.3390/sym14010117

[3] Makhdoom, Imran, Mehran Abolhasan, and Justin Lipman. "A comprehensive survey of covert communication techniques, limitations and future challenges." *Computers and Security* 120. (2022). https://doi.org/10.1016/j.cose.2022.102784

[4] Md Amiruzzaman. "A survey on steganography and steganalysis techniques in secret communication." *Research Briefs on Information and Communication Technology Evolution* 8, no. 7 (2022): 97–113. https://doi.org/10.56801/rebicte.v8i.139

[5] Bandyopadhyay, Samir Kumar, Vishal Goyal, Shawni Dutta, Sabyasachi Pramanik, and Hafiz Husnain Raza Sherazi. "Unseen to seen by digital steganography: Modern-day data-hiding techniques." In *Multidisciplinary Approach to Modern Digital Steganography*, pp. 1–28. IGI Global, 2021. https://doi.org/10.4018/978-1-7998-7160-6.ch001

[6] Tabares-Soto, Reinel, Raúl Ramos-Pollán, Gustavo Isaza, Simon Orozco-Arias, Mario Alejandro Bravo Ortíz, Harold Brayan Arteaga Arteaga, Alejandro Mora Rubio, and Jesus Alejandro Alzate Grisales. "Digital media steganalysis." In *Digital Media Steganography*, pp. 259-293. Academic Press, 2020. https://doi.org/10.1016/B978-0-12-819438-6.00020-7

[7] Michaylov, Kristiyan. "Exploring the use of steganography and steganalysis in forensic investigations for analysing digital evidence." Bachelor's thesis, University of Twente, 2023.

[8] Li, Li, Xinpeng Zhang, Kejiang Chen, Guorui Feng, Deyang Wu, and Weiming Zhang. "Image steganography and style transformation based on generative adversarial network." *Mathematics* 12, no. 4 (2024): 615. https://doi.org/https://doi.org/10.3390/math12040615

[9] Sayed, Mohamed H, and Talaat M Wahbi. 2024. "Information security for audio steganography using a phase coding method." *European Journal of Theoretical and Applied Sciences* 2, no. 1 (2024): 634–47. https://doi.org/10.59324/ejtas.2024.2(1).55

[10] Mao, Xueying, Xiaoxiao Hu, Wanli Peng, Zhenliang Gan, Qichao Ying, Zhenxing Qian, Sheng Li, and Xinpeng Zhang. "From covert hiding to visual editing: robust generative video steganography." *ArXiv Preprint ArXiv:2401.00652.* (2024). https://doi.org/https://doi.org/10.48550/arXiv.2401.00652

[11] Utama, Sunariya, and Roshidi Din. "Performance review of feature-based method in implementation text steganography approach." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 28, no. 2 (2022): 325-333. https://doi.org/10.37934/araset.28.2.325333

[12] Islam, Muhammad Aminul, Md Al Amin Khan Riad, and Tanmoy Sarkar Pias. "Performance analysis of steganography tools." In *2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, pp. 428–33. IEEE, 2020. https://doi.org/10.1109/ICAICT51780.2020.9333473

[13] Arora, Dr. Nitika. "Types and tools of steganography." *International Journal for Research in Applied Science and Engineering Technology* 10, no. 6 (2022): 2049–53. https://doi.org/10.22214/ijraset.2022.44279

[14] Bistarelli, Stefano, Michele Ceccarelli, Chiara Luchini, Ivan Mercanti, and Francesco Santini. "A survey of steganography tools at Layers 2-4 and HTTP." In *18th International Conference on Availability, Reliability and Security*, pp. 1–9. 2023. https://doi.org/10.1145/3600160.3605058.

[15] Reyers, Pieter Matthijs. "A comparative analysis of audio steganography methods and tools." Bachelor's thesis, University of Twente. 2023.

[16] Kheddar, Hamza, Mustapha Hemis, Yassine Himeur, David Megías, and Abbes Amira. "Deep learning for diverse data types steganalysis: A review." *ArXiv Preprint ArXiv:2308.04522.* 2023. https://doi.org/https://doi.org/10.48550/arXiv.2308.04522.

[17] Dehdar, Abouzar, Ahmad Keshavarz, and Naser Parhizgar. "Image steganalysis using modified graph clustering based ant colony optimization and Random Forest." *Multimedia Tools and Applications* 82, no. 5 (2023): 7401–18. https://doi.org/10.1007/s11042-022-13599-0

[18] Lerch-Hostalot, Daniel, and David Megias. "Real-world actor-based image steganalysis via classifier inconsistency detection." In *18th International Conference on Availability, Reliability and Security*, pp. 1–9. https://doi.org/10.1145/3600160.3605042

[19] Michaylov, Kristian D, and Dipti K Sarmah. "Steganography and steganalysis for digital image enhanced forensic analysis and recommendations." *Journal of Cyber Security Technology*, (2024). 1–27. https://doi.org/10.1080/23742917.2024.2304441

[20] Lee, Daewon, Tae Woo Oh, and Kibom Kim. "Deep audio steganalysis in time domain." In *ACM Workshop on Information Hiding and Multimedia Security*, pp. 11–21. 2020. https://doi.org/10.1145/3369412.3395064

[21] Martyniuk, Hanna, Valeriy Kozlovskiy, Tetiana Meleshko, and Anton Sorokun. "Method of finding cover signal for audio steganalysis calibrated methods." In *11th International Conference on Intelligent Data Acquisition and*

Advanced Computing Systems: Technology and Applications (IDAACS), 2 pp. 1095–1100. IEEE, 2021. https://doi.org/10.1109/IDAACS53288.2021.9661059

[22] Li, Jun, Minqing Zhang, Ke Niu, Yingnan Zhang, and Xiaoyuan Yang. "Motion vector-domain video steganalysis exploiting skipped macroblocks." *IET Image Processing*, (2023): 1–13. https://doi.org/10.1049/ipr2.13014

[23] Keizer, Mart, Zeno Geradts, and Meike Kombrink. "Forensic Video Steganalysis in Spatial Domain by Noise Residual Convolutional Neural Network." *arXiv preprint arXiv:2305.18070* (2023). https://doi.org/10.48550/arXiv.2305.18070

[24] Din, Roshidi, Azman Samsudin, and Puriwat Lertkrai. "A framework components for natural language steganalysis." International Journal of Computer Theory and Engineering 4, no. 4 (2012): 641.

[25] Din, Roshidi, and Azman Samsudin. 2009. "Intelligent steganalytic system : Application on natural language environment." *WSEAS Transactions on Systems and Control* 4, no. 8 (2009): 379–88.

[26] Huang, Huajun, Junshan Tan, Xingming Sun, and Lingxi Liu. "Detection of hidden information in tags of webpage based on tag-mismatch." In *3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP)*, 1 pp. 257–60. IEEE, 2007. https://doi.org/10.1007/978-3-642-04438-0_25

[27] Xinxin, Zhao, Huang Liusheng, Li Lingjun, Yang Wei, Chen Zhili, and Yu Zhenshan. "Steganalysis on character substitution using support vector machine." In *2nd International Workshop on Knowledge Discovery and Data Mining (WKKD)*, pp. 84–88. IEEE, 2009. https://doi.org/10.1109/WKDD.2009.105

[28] Din, Roshidi, Zhamri Che Ani, and Azman Samsudin. "A formulation of conditional states on steganalysis approach." *WSEAS Transactions on Mathematics* 11, no. 3 (2012): 173–82.

[29] Huang, Junwei, Xingming Sun, Huajun Huang, and Gang Luo. "Detection of hidden information in webpages based on randomness." In *3rd Internationl Symposium on Information Assurance and Security (IAS)*, pp. 447–52. IEEE, 2007. https://doi.org/10.1109/IAS.2007.74

[30] Huang, Huajun, Shaohong Zhong, and Xingming Sun. "Steganalysis of information hidden in webpage based on higher-order statistics." *International Symposium on Electronic Commerce and Security*, pp. 957–60. IEEE, 2008. https://doi.org/10.3724/SP.J.1146.2009.00530

[31] Xin-guang, Sui, Luo Hui, and Zhu Zhong-liang. "A steganalysis method based on the distribution of first letters of words." In *International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pp. 369–72. IEEE, 2006. https://doi.org/https://doi.org/10.1109/IIH-MSP.2006.265019

[32] Sui, Xin Guang, and Hui Luo. "A steganalysis method based on the distribution of space characters." In *International Conference on Communications, Circuits and Systems (ICCCAS)*, 1 pp. 54–56. IEEE, 2006. https://doi.org/10.1109/ICCCAS.2006.284584

[33] Chen, Zhili, Liusheng Huang, Haibo Miao, Wei Yang, and Peng Meng. "Steganalysis against substitution-based linguistic steganography based on context clusters." *Computers and Electrical Engineering* 37, no. 6 (2011): 1071–81. https://doi.org/10.1016/j.compeleceng.2011.07.004

[34] Chen, Zhili, Liusheng Huang, and Wei Yang. 2011. "Detection of substitution-based linguistic steganography by relative frequency analysis." *Digital Investigation* 8, no. 1 (2011): 68–77. https://doi.org/10.1016/j.diin.2011.03.001

[35] Meng, Peng, Liusheng Hang, Wei Yang, Zhili Chen, and Hu Zheng. "Linguistic steganography detection algorithm using statistical language model." In *International Conference on Information Technology and Computer Science (ITCS)*, 2 pp. 540–43. IEEE, 2009. https://doi.org/10.1109/ITCS.2009.246

[36] Yu, Zhenshan, Liusheng Huang, Zhili Chen, Lingjun Li, Xinxin Zhao, and Youwen Zhu. "Detection of synonym-substitution modified articles using context information." In *2nd International Conference on Future Generation Communication and Networking (FGCN)*, 1 pp. 134–39. IEEE, 2008. https://doi.org/10.1109/FGCN.2008.39

[37] Xiang, Lingyun, Xingming Sun, Gang Luo, and Can Gan. "Research on steganalysis for text steganography based on font format." In *3rd International Symposium on Information Assurance and Security (IAS)*, pp. 490–95. IEEE, 2007. https://doi.org/10.1109/IAS.2007.48

[38] Zhi-Li, Chen, Huang Liu-Sheng, Yu Zhen-Shan, Li Ling-Jun, and Yang Wei. "A statistical algorithm for linguistic steganography detection based on distribution of words." In *3rd International Conference on Availability, Security, and Reliability (ARES)*, pp. 558–63. IEEE, 2008. https://doi.org/10.1109/ARES.2008.61

[39] Chen, Zhili, Liusheng Huang, Zhenshan Yu, Wei Yang, Lingjun Li, Xueling Zheng, and Xinxin Zhao. "Linguistic steganography selection using statistical characteristics of correlations between words." In *International Workshop on Information Hiding*, pp. 224–35. 2008. https://doi.org/10.1007/978-3-540-88961-8-16

[40] Zhi-li, Chen, Huang Liu-sheng, Yu Zhen-shan, Zhao Xin-xin, and Zheng Xue-ling. "Effective linguistic steganography detection." In *8th International Conference on Computer and Information Technology Workshops*, pp. 224–29. IEEE, 2008. https://doi.org/10.1109/CIT.2008.Workshops.69

[41] Chen, Zhili, Liusheng Huang, Peng Meng, Wei Yang, and Haibo Miao. 2011. "Blind linguistic steganalysis against translation based steganography." In *International Workshop on Digital Watermarking*, pp. 251–65. https://doi.org/10.1007/978-3-642-18405-5-21

[42] Lingjun, Li, Huang Liusheng, Yang Wei, Zhao Xinxin, Yu Zhenshan, Chen Zhili, and C S Depart. "Detection of word shift steganography in PDF document." In *4th International Conference on Security and Privacy in Communication Networks*, pp. 22–25. https://doi.org/https://doi.org/10.1145/1460877.1460897

[43] Taskiran, Cuneyt M, Umut Topkara, Mercan Topkara, and Edward J Delp. "Attacks on lexical natural language steganography systems." In *Security, Steganography, and Watermarking of Multimedia Contents VIII*, 6072 pp. 97–105. SPIE, 2006. https://doi.org/10.1117/12.649551

[44] Meng, Peng, Liusheng Hang, Wei Yang, and Zhili Chen. "Attacks on translation based steganography." In *Youth Conference on Information, Computing and Telecommunication*, pp. 227-230. IEEE, 2009. https://doi.org/10.1017/CBO9781107415324.004

[45] Dalal, Mukesh, and Mamta Juneja. "Steganography and steganalysis (in digital forensics): A cybersecurity guide." Multimedia Tools and Applications 80, no. 4 (2021): 5723–71. https://doi.org/10.1007/s11042-020-09929-9

[46] Ali, Ahmed Mohamed Gaafar Mahmoud, Kelvin Chong Boon Kai, and Zool Hilmi Ismail. " Blockchain technology in overcoming security threats for smart manufacturing system - A systematic literature review." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 39, no. 1 (2024): 43-58. https://doi.org/0.37934/araset.39.1.4358

[47] Ingle, Darshan, and Divyanka Ingle. "An enhanced blockchain based security and attack detection using transformer in iot-cloud network." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 31, no. 2 (2023): 142-156. https://doi.org/10.37934/araset.31.2.142156

[48] Ali, Aoday H., Marwan B. Mohammed, and Dlsoz abdalkarim Rashid. "Classification of the Modern Approaches in Steganalysis Technique, Scenarios Attacks and Future Trends."