The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 179-184
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9487

Analysis of Data Security Resilience in Text Steganography on Indonesian Language Structure

R. Fanry Siahaan*, Dedi Candro P. Sinaga, Zanziqbar Alaydrus, Ikhwan Rafif Fadhil
Information Technology, STMIK Pelita Nusantara, Medan, Indonesia
Email: [1,*]rfanry@gmail.com,[2]dedisisnaga27@gmail.com, [3]alaydrus21@gmail.com,[4]ikhwan.rf@gmail.com
(* : rfanry@gmail.com)
Submitted: **16/11/2025**; Accepted: **29/11/2025**; Published: **30/11/2025**

Abstract— An in-depth analysis of data security in text-based steganography is necessary to ensure the sustainability and security of the methodology used. The purpose of this study is to analyze the resilience of data security in text-based steganography. The analytical approach used involves identifying and assessing the vulnerabilities of text steganography methods using Indonesian sentence patterns. The initial stage of the research was to analyze previous works related to this field to understand previously identified vulnerabilities. The applied text embedding model is based on a dictionary consisting of 1,929 words grouped into seven word categories that correspond to sentence patterns in Indonesian, namely adj (adjective), adv (adverb), nom (noun), num (numeral), par (particle), pro (pronoun), and ver (verb). Each word class is arranged into a sentence structure and each has the same bit length, namely eight bits. The robustness analysis results show that single-word input data is still vulnerable to brute-force attacks or pattern analysis if the message embedding process uses a simple sentence structure. This is due to the relatively small search space, which makes it easier for attackers to guess the embedding pattern. Conversely, using sentence patterns consisting of more than two words significantly increases combinatorial complexity and expands the possibility space, making hacking attempts much more computationally difficult. Thus, the robustness of a steganographic system increases as the number of words in the sentence pattern increases, as the time and resources required to perform the attack become practically inefficient.

Keywords: Data Protection; Text Steganography; Sentence Structure; Resistance to Attacks; Information Security.

1. INTRODUCTION

In data communication on a computer network, one of the crucial things that needs to be understood and paid attention to is information security. One important aspect in information security is integrity (authenticity), which ensures the authenticity of computer network user information so that the information cannot be changed by unauthorized parties [1], [2]. With the development of technology, all necessary information can be accessed easily, including confidential or very important ones. With the support of technology, all confidential information that is locked or stored properly can be accessed by irresponsible parties if the information security method used is simple or easy to guess [3] and [4]. Theft of user information such as usernames and passwords from an account or important data often occurs due to a lack of protection for confidentiality, integrity, and availability [5], [6]. In a computer network, this will certainly have serious consequences for the network's users. Therefore, an analysis is needed that aims to assess the level of confidentiality through the resilience of a message or user information in a computer network to improve the network in terms of protecting the confidentiality of user information [7], [8], [9].

Text-focused steganography is a technique for hiding hidden messages in plain-looking text. In the Indonesian context, data security assessments in steganography using text in Indonesian sentence structures are important because of the need to protect the confidentiality of information in a language spoken by the majority of the Indonesian population. The use of steganography with the LSB (Least Significant Bit) method to protect various text file data is quite robust against brute-force attacks [10], while the combination of the LSB method with the Redundant Count method in hiding messages is quite effective in deceiving steganological attacks [11]. Several studies on the use of steganographic media that are still ongoing until now show that the proposed Bayesian method shows superior performance compared to existing steganalysis methods for detecting various steganography in low-bit-rate compressed messages AbS-LPC [4] and [12]. Research on text-based steganography in the aspect of information hiding shows that without modification to the carrier text, it will increase the effectiveness of steganography, where the parity feature shows an even distribution in each character [3], [13].

In data security analysis, the main objective is to assess the extent to which text-based steganography is able to hide confidential information within Indonesian sentences without being detected by unauthorized parties and to ensure the continued security of the method. Specifically, robustness is evaluated through the exponential growth of the number of possible sentences generated by a dictionary containing 1,929 words grouped into seven word classes. Sentence patterns with one word produce a limited and relatively easy-to-analyze space of possibilities, while patterns with more than two words produce a much larger space of possibilities, making brute-force attacks practically inefficient.

This research presents a novel robustness-oriented analysis of text-based steganography by leveraging the syntactic structure of the Indonesian language as the primary security mechanism. Unlike most existing studies that focus on embedding techniques, payload capacity, or steganalysis detection accuracy, this work emphasizes resistance to brute-force attacks through combinatorial sentence space analysis. By utilizing a structured dictionary consisting of 1,929 Indonesian words categorized into seven grammatical classes, this study demonstrates that the number of possible syntactically valid sentences grows exponentially as sentence patterns increase in length. This exponential growth significantly enlarges the search space for unauthorized attackers, rendering brute-force attempts computationally

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9487

impractical without modifying the carrier text. Consequently, the proposed approach introduces a language-driven robustness metric for text steganography in Indonesian, offering a distinct contribution to information hiding research by integrating linguistic variability with security evaluation.

2. RESEARCH METHODOLOGY

2.1 Research Stages

In this study, the resilience of text steganography is defined as the ability of the system to increase the difficulty for attackers to extract hidden messages without knowing the sentence patterns and dictionary structure used. Resilience is analyzed using several indicators, namely: first, the size of the search space resulting from the combination of word classes and sentence pattern length; second, the combinatorial complexity of sentence patterns determined by the number of words in a sentence structure; and third, the level of vulnerability to brute-force attacks and linguistic-based pattern analysis. The larger the search space and sentence pattern complexity, the higher the system's resilience because the time and computational resources required to carry out an attack increase significantly. Thus, an increase in the number of words in a sentence pattern contributes directly to an increase in the resilience of the proposed text steganography. The research method applied in this study is descriptive qualitative where qualitative data consists of descriptive data in the form of numerical symbols. Qualitative data is carried out to understand empirical phenomena with the aim of finding as many descriptions as possible without detailing the relationship between variables, which is carried out through several stages in the following figure

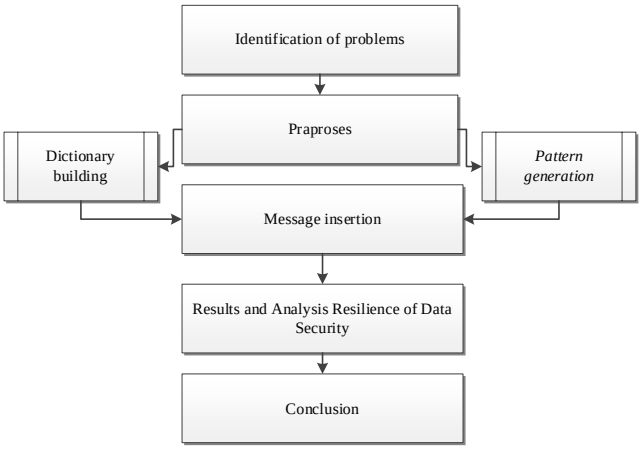

Fig 1. Methodology Research

2.2 Information Collection

The method applied in data collection is descriptive qualitative, with data sources coming from the 2008 Big Indonesian Dictionary, which consists of 7 types of words. In the text, the citation numbers are in sequence in square brackets [3], also tables of numbers and numbers in sequence as shown in Table 1 and Figure 1.

Table 1. Types of words

Abbreviation	Word Class	Total Word Dictionary	Description
Ver	Verb	368	Verb
Pro	Pronoun	4	Pronouns, demonstratives and question words
Par	Particle	41	Conjunctions
Num	Numerals	5	Number words
Name	Noun	1,265	Noun
Adv	Adverbial	14	Adverb
Adj	Adjective	232	The word sigat

2.3 Prepocessing

The preprocessing stage is the data filtering or selection stage that aims to obtain accurate data that will be used in the next process. In this preprocessing stage, two input components are formed, namely a dictionary and a pattern. These components, along with a secret message, will function as input. A hidden message is added. The cover media is the location for placing the message generated from the dictionary and pattern elements during the insertion process. The number of words for each word type (f) is limited by a power of two with the calculation g = ⌊2^log_2⌋f ⌋ The sentence structure used is a combination of the grammatical functions of sentences in Indonesian, namely subject (S), predicate (P), object (O), complement (Pel), and adverb (K). From these five grammatical functions, sentence patterns in Indonesian

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9487

can be derived [11], [13]. Patterns are formed by recording all possible combinations (cross products) of the grammatical elements of 8 basic sentence patterns. Each combination of grammatical elements in each pattern is stored in a pattern file. To solve the problem of limited message length, the pattern uses the function SIZERR(C) → C¯ + C + R In this article, only three sentence patterns are applied, as shown in the following table.

Table 2. Types of Sentence Pattern

Sentence Pattern	Style Contents
SP	nom-num
SP-Pel	nom-ver-num
SPO-Pel	pro-adj-nom-num

Stegotext is a combination of a dictionary and a sentence structure that serves to hide secret information. The stegotext created by SIZER consists of a sequence of a certain length representing the length C (Cbit), followed by a binary string of the message (C), then added again with a random string R. This is the most crucial component in this research, which will be evaluated, to what extent the setegotext is resistant to distortion attacks via bruteforce (testing all possibilities).

2.3 Information Prepocessing

The components that need to be considered in generating a random string R of r bits in the SIZER process are p (message length), s (the number of bits that can be hidden in the pattern), n (the length of the bits that represent the length of the message), which is the maximum amount of data in the dictionary so that n can be represented by a maximum of 2n bits of binary message, and r (the number of random bits generated). Therefore, if $p + n \neq s$, then $r = (s * x) - (p + n)$, where x is the pattern repetition variable that determines how many times a sentence with the pattern used will be generated so that all message bits can be transformed, thus forming a stegotext. The original message that will be used as sample data in analyzing resistance to attacks in the form of distortion from external parties as shown in table 3 below.

Table 3. Sample data in analyzing

No.	Message	Information		
		(say)	(character)	(bit)
1	A tertiary institution is an educational institution that provides higher education and can take the form of an academy, polytechnic, college, institute or university.	19	166	1,328
2	A lecturer's primary responsibility is to conduct education, research, and community service. Furthermore, a lecturer is expected to be able to plan and implement the teaching process, as well as assess and evaluate learning outcomes.	29	253	2,024

3. RESULT AND DISCUSSION

Based on the results of the proposed text steganography resilience analysis, the main findings of this study can be summarized as follows:

1. One-word sentence patterns produce a limited space of possibilities, making them vulnerable to brute-force attacks and pattern analysis.
2. Two-word sentence patterns significantly increase combinatorial complexity, but analysis is still possible with certain computational resources.
3. Sentence patterns with more than two words result in exponential growth in the space of possibilities, making hacking attempts practically inefficient.
4. The use of classified dictionaries with uniform bit lengths maintains the consistency of data insertion without reducing the linguistic structure of the sentence.

3.1 SIZER Procedure

The process of transforming the original message into stegotext which will then be analyzed for its resistance to external interference is by representing the message length (n) obtained from (Total Dictionary words)/log$_{f_0}$2, namely by rounding up, so that the length $n=\lceil 1929/\log_{f_0}2 \rceil=11$ bits.

a. Pattern number two for data sample number one

$x = \lceil (1.328+11)/24 \rceil = \lceil 1339/24 \rceil \approx 56$, the second pattern will be used 56 times to produce the stegotext.

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9487

Stegotext:
various anachronisms. I read anomalies. I invite alternatives. I am good. I scatter threats. I am neglected by the ark. I am anti-bahara. I am responsible for the heat. I am caring for the anghur. I adjust my program. I reach ablaif. I am good at alliteration. I teach paragraphs. I take athletes. I am fencing algebra. I am visiting advertising works. I divide the water. I am windy in nature. I am involved in astronomy. I am raising squirrels. I am breaking up the program. I am clashing with alluvium. I stir up the affect. I am antidote. I am wearing out my father. I am spreading angklung. I am souring hope. I am caring for alai-belai. I am reading alkenes. I am good at virtue. I am organizing activists. I am reading acclamation. I am hijacking andiko. asta acung ensemble. ampat menganyang antidioxide. ampat mengangin-airkan atok. ampat beradu adik. ampat acau adad. ampat membaham affiliisasi. asta ambur arut. ampat directing archives. asta mengari adik. asta ayun-temayun alkana. asta beranjangkarya adai-badai. berbagai beranggul arput. berbagai angglap azam. ampat float familiar. diverse lift amril. diverse berserk archives. ampat articulate agut-agut. diverse atung sickle. asta pembagul apu-apu. diverse amputir akan. asta membabak badong. aku mengarangang angkan-angkan. asta beracara bahara.

b. The third pattern for data sample number one

x = ⌈(1.328+11)/32⌉= ⌈1339/32⌉ ≈42, the third pattern will be used 42 times to produce the stegotext.
Stegotext:
badang various kinds of absorption are diverse. badang good headed four. why aggressive spirit four. why associative grass four. badang calm arung four. why affective view me. is wrong current four. why Almasih environment four. badang antagonist hobby me. badang sterility absorption is diverse. why natural merger me. why doubt adventurer foot. badang positive commander foot. why pity my contract. badang various kinds of adenoma are diverse. is good bestowal foot. badang less antagonistic me. is gray essence me. why adiabatic intestine foot. why critical worm medicine four. badang combination of drugs four. is good colored me. why graceful fetus four. badang big amra foot. why babil antifreeze four. badang additional agar-agar four. why quick mind me. is silly belief four. badang hard mercy foot. badang bacek adipati diverse. badang budget award me. badang red norm is diverse. why good agromania four. is analytical argentometer foot. why beauty luxury four. badang absolute plan foot. why is erosion hairy? adiabatic body? emotion? is there a difference.

c. Pattern number two for data sample number two

x = ⌈(2.024 + 11)/24⌉ = ⌈2035/24⌉ ≈ 79, the second pattern will be used 79 times to produce the stegotext.
Stegotext:

ampat clarifies adenoma. asta invites acerang. ampat messes up aerometer. ampat warns asai. asta cooperates with amuk. ampat reduces asabat. ampat regulates adrenal. ampat supports analeptic. ampat responds to threats. various awaits aurora. asta fantasizes anyang. various tells a story. ampat validates somewhat. asta flows current. ampat imagines agglutinogen. ampat imagines ablepsia. various describes akasia. ampat is agitated somewhat. ampat threatens agar-agar. various grants alliance. i tear up the bible. asta nods anaphora. aneka ignores anesthesia. asta plans aristocrat. asta chases almuazam. ampat windy advocate. ampat focuses part. ampat fantasizes avgas. ampat arsenic validation. ampat collects anemokori. asta directs anesthesia. various inter-ampere. asta improves anduhan. i influence argument. Asta moves on. Asta centuries alif. Ampat rises aji. Ampat realizes adagium. Ampat parades angsana. Various avocados. Ampat considers teachings. Asta mixes water. Ampat records abadiah. Ampat burns anthrax. Ampat resonates alamas. Various have accusative members. Various deliver transmitters. Asta rejects astringents. Ampat discusses alveoli. Asta discusses andeng-andeng. Asta complains about swans. Ampat arranges ayuman. Various monitors arput. Various nods the roof. Various lifts arem-arem. Various mixes adverbs. I recommend acang. Ampat arranges b. Various explains administrator. I consider aversion. Ampat nods adication. Asta feels ameiosis. I build anghur. Ampat cloudy alabio. Ampat falls anghur. Asta argues amor. Ampat has an introduction. Various adapt amphibians. Various imagine clothes. Various forgive alang-alang. asta adopts the article. i plan autarky. asta disrupts the lesson. various altruistic ayum. various attempts at wear and tear. ampat passes through addiction. asta airs out asar. various antiseptic mengarun. various playing ara. asta details andur. i love atma. ampat organizes origin.

d. The third pattern for data sample number two

x = ⌈(2024+11)/32⌉= ⌈2035/32⌉ ≈64, the third pattern will be used 64 times to produce the stegotext.
stegotext:
gray matter alphabet four. avirulent aerosol matter diverse. why alaihiasalam antarctica four. is anonymous anticadar diverse. agrometeorological matter ambring-ambringan matter diverse. why ambring-ambringan antidioxide four. ambivalent aggression matter four. why apatite asepsis diverse. why antivenom analogs diverse. adaptive matter if i. why shame aedes four. why automatic anchor diverse. allegorical matter adiasam diverse. anonymous matter ammunition asta. why adequate proverb four. why archaic antan four. calm matter agal four. why babar ammunition asta. calm matter envelope asta. why angker alkalimerkurium four. why awahama arai asta. why ala pharmacist diverse. anom antiasam four. asak badang anthroponymy asta. random body aviator asta. is acrophobia archaeologist asta. is the alantois asta mess. badang asih aksep four. badang ammi agenda diverse. is aprit alantoikase four. is adn agroecosystem diverse. badang gray alinasi aku. is gray alkana four. badang aditional civilization four. badang ambekparamarta agar-agar four. why gray audiophone asta. why directions abaian diverse. badang abasah aquarium four. badang apik aristocracy asta.

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9487

why additional antacid four. is acrophobia ajufan diverse. is asor aliphatic asta. badang baik apit four. why aboral anus four. is adiabatic alternative four. badang aggressive bahana four. badang arau abadi aku. badang angah aliphatic diverse. why anthesis afdal four. is accurate adenoma four. badang alap ayut-ayutan four. why aus ajar asta. badang ampuh alkisah four. is alot pengawam asta. badang ayu suhung variasi. badang accelerometer adai-badai four. why is the agrarian smell asta. badang laughing athlete four. why is the autopsy akmal four. why is the antique abilah asta. badang alah badur asta. why is the salty antelas four. why is the bacak apelativa asta. badang apes alinea four.

3.2 Resilience Analysis

Referring to the SIZER results above, namely the number of words per class, the total for the first pattern is 237 words, for the second pattern the total is 1,638 words and the total for the third pattern is 1,506 words. By utilizing the formula to calculate the combination probability, namely P(n,r)=n!/(nr)!. Assessing the resistance to attacks in revealing the message content of each entered text is as follows:
Second pattern = 1638!/((1638-19)!)= 1638!/1619!=infinity
The third pattern = 1506!/((1506-19)!) = 1506!/1487! = infinity
The third input text consists of 30 words.
The first pattern (n,r)=n!/(nr)!=237!/(237-30)!=237!/207!=infinity
Second pattern = 1638!/((1638-30)!) = 1638!/1608! = infinity

Table 4. Resilience Analysis

Style Contents	Input text		Stegotext		Endurance Analysis	
	To-	Many Words	To-	Many Words	$P(n,r)=\frac{n!}{(n-r)!}$	Information
nom-num	1	1	1	12	237	Strong Enough
	2	19	2	12	infinity	Very strong
	3	30	3	12	infinity	Very strong
nom-ver-num	1	1	1	162	1,638	Strong
	2	19	2	168	infinity	Very strong
	3	30	3	159	infinity	Very strong
pro-adj-nom-num	1	1	1	257	1,506	Strong
	2	19	2	246	infinity	Very strong
	3	30	3	256	infinity	Very strong

4. CONCLUSION

The durability analysis results in Table 4 above show that input with a single word can still be hacked if using a short sentence pattern. Meanwhile, using sentence patterns consisting of more than two words becomes increasingly difficult and even impossible to hack due to the vast, even infinite, number of possibilities. The methodological contributions of this research include: first, the development of a text steganography model that utilizes Indonesian sentence structure as a medium for message insertion; second, the use of a classified dictionary consisting of seven word classes with uniform bit lengths to maintain data representation consistency; and third, an analytical approach in evaluating text steganography resilience based on linguistic complexity and possibility space, not only on insertion capacity. This approach provides a more comprehensive resilience evaluation framework against brute-force attacks and pattern analysis.

REFERENCES

[1] R. Gurunath, A. H. Alahmadi, D. Samanta, M. Z. Khan, and A. Alahmadi, "A Novel Approach for Linguistic Steganography Evaluation Based on Artificial Neural Networks," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3108183.
[2] M. A. Hameed, O. A. Abdel-Aleem, and M. Hassaballah, "A secure data hiding approach based on least-significant-bit and nature-inspired optimization techniques," *J. Ambient Intell. Humaniz. Comput.*, 2023, doi: 10.1007/s12652-022-04366-y.
[3] H. Huanhuan, Z. Xin, Z. Weiming, and Y. Nenghai, "Adaptive Text Steganography by Exploring Statistical and Linguistical Distortion," 2017. doi: 10.1109/DSC.2017.16.
[4] H. Kang, H. Wu, and X. Zhang, "Generative text steganography based on LSTM network and attention mechanism with keywords," 2020. doi: 10.2352/ISSN.2470-1173.2020.4.MWSF-291.
[5] L. Li, W. Zhang, C. Qin, K. Chen, W. Zhou, and N. Yu, "Adversarial batch image steganography against CNN-based pooled steganalysis," *Signal Processing*, 2021, doi: 10.1016/j.sigpro.2020.107920.
[6] S. Mukherjee, S. Mukhopadhyay, and S. Sarkar, "ChatGPT Based Image Steganography (CGIS): A Novel Intelligent Information Hiding Approach to Achieve Secure Covert Communication," 2023. doi: 10.1109/ICAEECI58247.2023.10370937.
[7] W. Peng, S. Li, Z. Qian, and X. Zhang, "Text Steganalysis Based on Hierarchical Supervised Learning and Dual Attention Mechanism," *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2023, doi: 10.1109/TASLP.2023.3319975.

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9487

[8] W. Su, J. Ni, X. Hu, and J. Fridrich, "Image Steganography with Symmetric Embedding Using Gaussian Markov Random Field Model," *IEEE Trans. Circuits Syst. Video Technol.*, 2021, doi: 10.1109/TCSVT.2020.3001122.

[9] Y. Tong, Y. L. Liu, J. Wang, and G. Xin, "Text steganography on RNN-Generated lyrics," *Math. Biosci. Eng.*, 2019, doi: 10.3934/mbe.2019271.

[10] X. Wang, Y. Wang, K. Chen, J. Ding, W. Zhang, and N. Yu, "ICStega: Image Captioning-based Semantically Controllable Linguistic Steganography," 2023. doi: 10.1109/ICASSP49357.2023.10095722.

[11] Z. L. Yang, S. Y. Zhang, Y. T. Hu, Z. W. Hu, and Y. F. Huang, "Vae-stega: Linguistic steganography based on variational auto-encoder," *IEEE Trans. Inf. Forensics Secur.*, 2021, doi: 10.1109/TIFS.2020.3023279.

[12] S. Zhang, Z. Yang, J. Yang, and Y. Huang, "Provably Secure Generative Linguistic Steganography," 2021. doi: 10.18653/v1/2021.findings-acl.268.

[13] X. Zhou, W. Peng, B. Yang, J. Wen, Y. Xue, and P. Zhong, "Linguistic Steganography Based on Adaptive Probability Distribution," *IEEE Trans. Dependable Secur. Comput.*, 2022, doi: 10.1109/TDSC.2021.3079957.