Analysis Review on Performance Metrics for Extraction Schemes in Text Steganography

Raihan Sabirah Sabri[1], Roshidi Dini[2], Aida Mustapha[3]
[1,2]School of Computing UUM College Arts and Sciences, Universiti Utara Malaysia, 06010, Sintok, Kedah, Malaysia
[3]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia,
Parit Raja, 86400 Batu Pahat, Johor, Malaysia

Article Info

Article history:

Received Apr 1, 2018
Revised Jun 20, 2018
Accepted Jul 7, 2018

Keywords:

Cryptography
Extracting process
Performance
Steganography

ABSTRACT

In today's era of big data computer networks, protection of secret messages when transmitting information is a major concern. The openness and publicity of the communication channel are the main attraction for malicious people to steal personal data even though privacy protection in operational. Data extraction is process that reverses the data embedding process in information hiding. However, the performance of an information hiding framework highly depends on the evaluation metrics used. The effectiveness of evaluation itself is mainly determined by the performance aspect or critera such as capacity, imperceptibility or security. The aim of this paper is to present a review on trends for existing performance metrics used in extraction schemes from a data hiding framework. This review is hoped to help future research in evaluating the performance of data hiding framework in general and the proposed extraction schemes in specific.

Corresponding Author:

Roshidi Dini,
School of Computing UUM College Arts and Sciences, Universiti Utara Malaysia,
06010, Sintok, Kedah, Malaysia.
Email: roshidi@uum.edu.my

1. INTRODUCTION

The development of the Internet and mobile communication requires a specialized security protocol especially when dealing with wide computers network. Network security is increasingly important due to the number of data being exchanged over the Internet [1]. Therefore, confidentiality and integrity of the data must be treated with the utmost importance in order to protect against unauthorized access. As mention in [2] the current situation in term of communication network still not properly secured, because communication content might be eavesdropped by third party besides identities could be pretend to be. Cause of that reason people become doubtful about communication network condition.

Basically, in data transmision the most important is security parameter [3]. One of technique to secure data is with steganography and cryptography [3]. A study by [4] stated the requirement performance metric of cryptography technique there shoud be have such as confidentiality, authentication, data integrity, non-repudiation, and service reliability and availability to produce a enhanced protected technique.

One way to secure the information exchange via the Internet is by means of data hiding. There are three methods generally used in data hiding consisted of cryptography, steganography and watermarking [5]. Cryptography is the process of message transfer in achieving confidentiality. Cryptography is known as secret writing to keep from the various attackers from stealing the secret information [6]. By using cryptography, the secret message will be converted into a cipher text in the form of plain text [6] .

While cryptography is the process of converting the message to be communicated into unintelligible messages, steganography involves the process of hiding the information within another cover medium like text, image, audio, video files [7]. Steganography provides better security protection of the data as compared to cryptography. Cryptography only aims to conceal the content of the secret message while steganography aims to conceal the contents of the secret message as well as the fact that a secret message is being concealed [7]. The advantage of steganography is that the hidden messages would not attract any attention because it is concealed in a way that it appears as part of the original medium.

Finally, watermarking is defined as the process of embedding watermarks in digital media such as audio, video and image. In digital watermarking, the watermark either a short text or a company logo is embedded into the image file [8]. This is a robust way such that any changes to the image, like compression, filtering, noise addition does not alter the quality of the image. A general data hiding framework is shown in Figure 1, where it consists of two main processes; embedding and extracting.

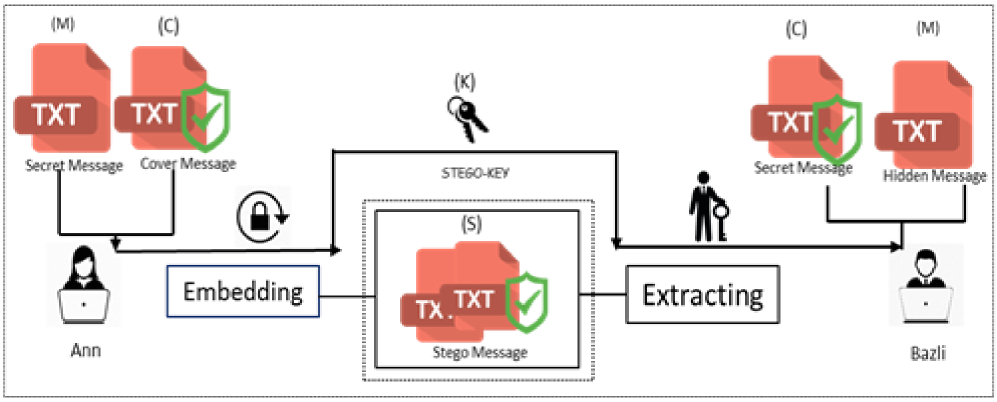

Figure 1. Data hiding framework

The process of embedding and extracting is illustrated in Figure 1 via two personnel; Ann and Bazli. To complete the process, Ann will embed the secret message (M) into the cover message (C) using stego-key (K) and produces the stego message before transmit to Bazli. Meanwhile, Bazli needs to use recovering algorithm to interpret with extracting process and obtain the secret message (M).

The data extraction process essentially converts the cipher text into the plain text by using a secret key. As an example, [9] mapped the pixels into an image, which is a type of extraction method. The requirements for data extraction in watermarking method include the watermarked image and the secret key. For extraction, the process applies the inverse of support vector dimension on watermarked image using a secret key.

The literature has shown a substantial contribution to different extraction techniques since the last decade. All the methods are evaluated based on different performance evaluation metrics, making a central comparison difficult even though they share the same goal in data hiding, which is to improve security, reliability, and efficiency.

Performances analysis is essential to measure the efficiency and effectiveness of a proposed extraction technique. Nonetheless, they are more aspects to measure rather than efficiency and effectiveness alone. The evaluation aspects must be determined in order to identify which performance aspect would be of emphasis in a particular research. Other than efficiency and effectiveness, robustness is also a type of popular aspect used as evaluation metric. Robustness measures the ability to survive intentional or unintentional attacks aiming at removing or modifying the embedded payload. Other example includes embedding distortion rate as the performance metric in the proposed data hiding schemes [10].

This paper is set to review different aspects used as performance metrics in evaluating extraction scheme as shown in Figure 1 in order to see research patterns for last 10 years. The remaining of this paper is organized as follows. The following Section 2 summarizes a list of extraction performance criteria with symbols representation. Section 3 presents the methodology used to analyze of the research trends in evaluation of extraction schemes. Section 4 delivers the finding derived from the previous sections and finally Section 5 will conclude research contribution in this paper.

2. REVIEW CRITERIA

There are various type of extraction performance criteria frequently used in data hiding. Table 1 shows the criteria on extraction performance types with the respective symbols used in the literature.

Table. 1 Classification Criteria on Performances Metric Representation

Criteria	Performance Metrics	Symbol
Extracton Scheme	Security [9], [11]–[19]	S
	Imperceptibility [11], [20]	IM
	Capacity [11]–[13][21]–[23]	C
	Robust [9], [19], [24]	R
	Efficiency [16], [18], [25]	E
	Invisibility [26]	IN
	Accuracy [27], [28]	AC
	Speed [15]	SP
	Complexity [21]	CP
	Reversibility [25], [29]	RV
	Realitibility	RL
	Integrity [24], [30]	IG
	Perceptibility [11]	PC
	Assumption [24]	AS
	Confidentiality [31]	CF
	Availability [32]	V

There are 16 performances available listed as a result of relevant search studies, for example security, imperceptibility, capacity, robustness, efficiency, invisibility, accuracy, speed, complexity, reversibility, realitibility, integrity, perceptibility, assumption, confidentiality, and availability. According to the observation, previous researchers attempt to improve or measuring data hiding performance around these aspects and focus on only one performance dimension or more. Meanwhile, in certain studies, researchers will use several aspects depending on requirement of approach to evaluate their particular goal such as approach techniques, algorithm and schemes to improve the performance. Time is another dimension that can be used to measure data hiding performance.

3. METHODOLOGY

Based on the performance metrics identified in Table 1, the resuts are analyzed to answer the following research questions:

a. What is the research trend on application of performance metrics in evaluating data extraction schemes?
b. What are the preferred or widely used performance metrics in evaluating data extraction schemes?
c. What is the top three performance metrics used in evaluating data extraction schemes?

The methodology used to answer the research questions in analyzing the performance of evaluation metrics for extraction schemes is shown in Figure 2.

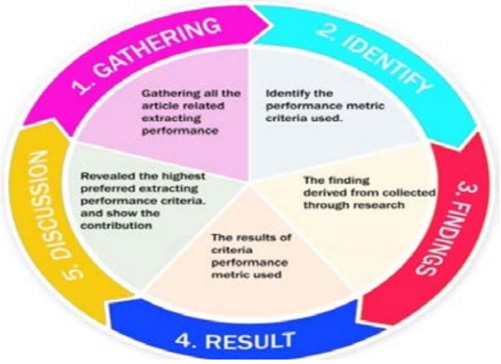
1. GATHERING
Gathering all the article related extracting performance
2. IDENTIFY
Identify the performance metric criteria used.
3. FINDINGS
The finding derived from collected through research
4. RESULT
The results of criteria performance metric used
5. DISCUSSION
Revealed the highest preferred extracting performance criteria. and show the contribution

Figure 2. Research methodology

From the figure, gathering is the process of collecting information through literature review and listing parameter criteria used by previous study. Next is the process of identifying the performance metrics or criteria used in previous research. Once the information is ready, the next step if deriving the findings into results that are represented in the form of percentage and figures. The final step is presenting the discussion from the review in order to find the most popular or most widely used performance criteria within the last decade together with analysis on the year that research on extraction schemes are most aggressive.

4. RESULTS AND ANALYSIS

This section presents the trends of extraction performance metrics used in previous studies. The first trend is based on the most frequently used extraction performance metrics, the second trend is to demonstrate the top three highest performance metrics used, and the final trend is the year when research in extraction schemes is most aggressive.

4.1. Trends of Performance Metrics Used in Extraction

Figure 3 shows the number of literature collected throughout the year 2002 until 2017. The most noticeable trend in Figure 3 is that capacity has been consistently measured in extraction schemes 2002. The research that focuses on measuring capacity peaked in 2010, 2013 and 2015. The capacity aspect concerns on the amount of hidden bytes over the size of the cover text in bytes. [33] Specified 4 parameter necessity for strong steganographic technique, one of them is capapcity parameter is the requirement "should be high"

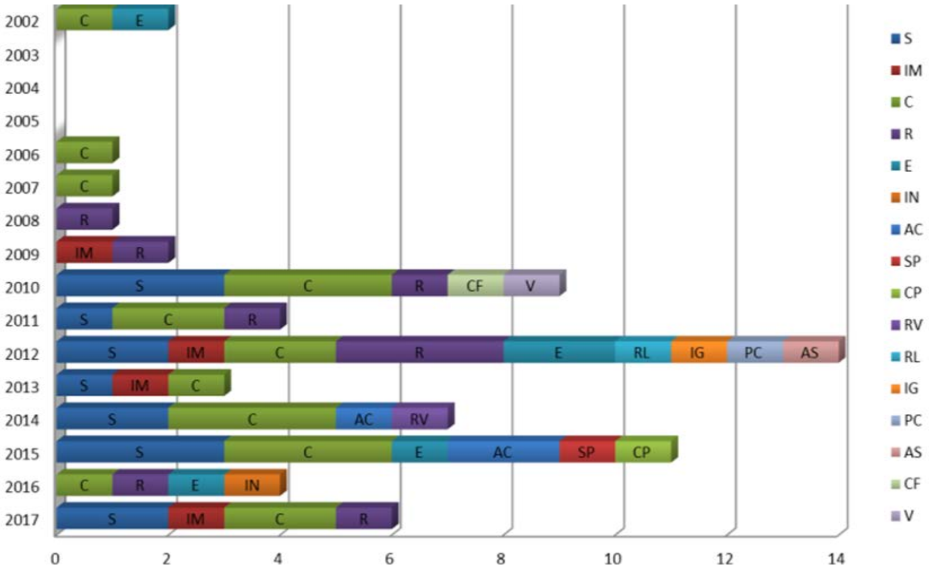

Figure 3. Trends of performance metrics in extraction schemes

Next, the security aspect was also consistently measured in 2010. This aspect focuses on the benefit of a proposed extraction schece from the security point of view [28]. A study by [32] stated that it is important to ensure that a proposed technique be able to provide double security or the capacity to embed more secret messages in order to increase the capability to exchange secret information. A good hidden invisibility is achieved with good security and good strength for a plenty of hidden attacks. While [33] revealed the parameter of security for good technique in steganography "should be high" same as capacity too.

Finally, the measurement on robustness of extraction schemes in data hiding was only started in 2008 with a small amount of studies. The trend maintains the same number of studies throughout the years but beginning to gain popularity in 2011. One particular importance of the robustness aspect is that while modifying the cover medium, the proposed techniques must not affect the process of embedding information during data hiding [31] mentioned that image-processing technique should provide robustness in stego-image when occurrence such as cropping and compression takes place. When any of this technique is executed, the hidden message must not completely damaged.

Figure 3 also demonstrates that the introduction on different performance metrics in extraction schemes was most active in 2012 with 14 contributions of study findings. However, the following three years had seen a decline in research with 11 studies in 2015 and down to 9 studies in 2010.

4.2. Preferred Extraction Performance Metrics

From the 16 performance metrics or aspects listed, Figure 4 shows that there are gaps of differences between each particular aspect whereby there exist three groups or patterns; widely used performance metrics, average use, and least used performance metrics. The group consisting the leading performance metrics include capacity (C), security (S) and robustness ® with 32%, 19%, and 14%, respectively. The second group consists of AC (5%), IM (6%) and E (8%). The remaining aspects in the last group are IN, SP, CP, RV, RL, IG, PC, AS, CF and other performance metrics which is the percentage classified smaller than 5%.

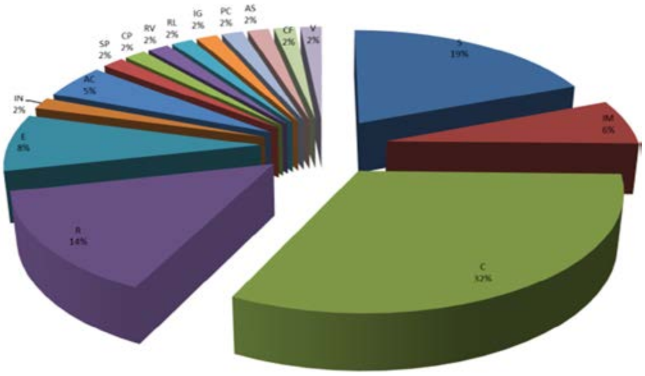

Figure 4. Preferred performance metrics

Overall, the majority of performance type preferred refers to the capacity (C) aspect about 32%. Generally, this aspect is about amount or volume of data that can be embedded in the cover medium or the capability of a cover media to hide secret information [9].

Second preferred metric obtain by security performance metric with 19% difference about 13% less. Security is about to hiding the real existence of communication through deceptive closures without notice with intruder [34].

Following capacity is robustness (R) with 14% contributions in the literature. Robustness concerns on modifying the stego object but not affecting the embedded information [24]. A study by [34] stated that in cryptography technique the robustness is the parameter to measuring the secrecy of data transmission, while security is measured the way silent of communicate data.

Finally, performance metrics such as SP, CP, RV, IG, PC, AS, CF and V are rarely used in evaluating extraction process during data hiding.

4.3. Top Three Extraction Performance Metrics from the Literature

Table 2 present the result of three highest preferred performance types from the previous studies. The percentage obtained are based on analysis in Section 4.2 where the top three performance metrics are diminished into a majority of focused form. This three highly rated performance metrics are separated from the 16 performances type available in Table 1.

Table 2. Percentage Use of Performance Type

Preferred Performance Type	Percentage (%)
Capacity	39
Security	37
Robust	24

Based on Table 2, capacity refers to the amount of secret information that can be embedded without affecting the carrier for instance medium. Among the preferred performance type list, capacity appeared as the top-most performance metrics used as compared to other performance criteria with 39%. This findings is supported by [35] who stated that a higher embedding capacity of any proposed data hiding technique will provide higher security and able to avoids suspicion on the existence of a secret message. Nonetheless, increasing capacity can affect the security performance [36]. For example, by increasing the size of sub-block size, the embedding capacity will increase but the security performance of the encryption algorithm will decrease.

Next, security acquired 37% from the majortity usage indicating that this is common aspect to evaluate in extraction schemes. Research by [17] highlighted the importance to consider potential attacks on security features and demonstrated that application of an XOR operation is able to increase the strength of any steganographic method and achieve a high level of security. Finally, robustness refers to the degree of difficulty required to destroy embedded information without destroying the cover medium. This metrics acquired 24% from the preference pool. Robustness was used in [32] to identify operational efficiency of the proposed steganographic system.

5. DISCUSSIONS AND CONCLUSIONS

In conclusion, since capacity is the most popular aspect measured for evaluation of an extraction scheme in a data hiding framework as shown in Figure 1, it is important to understand that hidden message capacity is highly relative to the size of the cover medium (i.e. image/text), which is with capacity aspect can provide better storage capacity. Figure 5 shows the three most preferred performance metrics, which are capacity aspect (32%), security aspect (19%), and robustness aspect (14%). The analysis also reveled that research on extraction schemes was most aggressive in the year of 2012 where 14 contributions were made during that year.

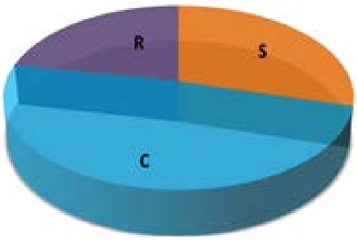

Figure 5. Three highly preferred performance metric

This review is hoped to help future research in evaluating the performance of data hiding framework in general and the proposed extraction schemes in specific.

ACKNOWLEDEMENT

This research is jointly supported by the Fundamental Research Grant Scheme (FRGS), MoHE under SO Code 13576 Universiti Utara Malaysia and Vot 1557 Universiti Tun Hussein Onn Malaysia.

REFERENCES

[1] A. Achuthshankar, “A Novel Symmetric Cryptography Algorithm for Fast and Secure Encryption,” pp. 4–9, 2015.
[2] H. Tu, “A Security Enhanced Password Authentication and Update Scheme Based on Elliptic Curve Cryptography,” vol. 12, no. 10, pp. 7353–7360, 2014.
[3] W. S. Sari, E. H. Rachmawanto, D. R. Ignatius, M. Setiadi, and C. A. Sari, “A Good Performance OTP Encryption Image based on DCT-DWT Steganography,” vol. 15, no. 4, pp. 1987–1995, 2017.
[4] C. Science, A. Academy, M. Saad, K. Youssef, and H. Abdel-kader, “Quantitative Analysis and Comparison of Symmetric Cryptographic Security Algorithms,” vol. 4, no. 1, pp. 116–124, 2016.
[5] F. M. Shelke, A. A. Dongre, and P. D. Soni, “Comparison of different techniques for Steganography in images,” *Int. J. Appl. or Innov. Eng. Manag.*, vol. 3, no. 2, pp. 171–176, 2014.
[6] A. Sheshasaayee, “Analysis of Techniques Involving Data Hiding and Watermarking,” no. Icimia, pp. 593–596, 2017.
[7] U. Terms and C. C. Concepts, Domain 3: Cryptography. 2010.
[8] R. Chandramouli, “Digital Watermarking,” no. 1, pp. 1–21.
[9] S. Chaudhary and M. Dave, “An Elucidation on Steganography and Cryptography,” pp. 3–8, 2016.
[10] S. Dey and A. Abraham, “Data Hiding Techniques Using Prime and Natural Numbers,” pp. 1–45.
[11] L. Kothari, “Data hiding on web using combination of Steganography and Cryptography,” pp. 448–452, 2017.
[12] H. Huanhuan, Z. Xin, Z. Weiming, and Y. Nenghai, “Adaptive Text Steganography by Exploring Statistical and Linguistical Distortion,” 2017.
[13] M. Y. Elmahi and M. H. Sayed, “Text Steganography Using Compression and Random Number Generators,” vol. 6, no. 6, pp. 259–263, 2017.
[14] A. Hamarsheh, “Exploiting Omega Networks to Hide Text-in-Text Messages,” vol. 15, no. 5, pp. 39–43, 2015.
[15] A. P. Kaur and G. Singh, “Designing and Performance Evaluation of Text Data Hiding Technique Using Sequential Encoding and Decoding Technique,” *Int. J. Adv. Res. Eng. Technol.*, vol. 3, no. Ii, pp. 6–11, 2015.

[16] P. M. Vidhya and V. Paul, *“A method for text steganography using malayalam text,”* Procedia Comput. Sci., vol. 46, no. Icict 2014, pp. 524–531, 2015.

[17] N. Elyatawfiq, “Hiding Text within Image Using LSB Replacement,” vol. 13, no. 3, pp. 13–16, 2013.

[18] R. F. Mansour, W. F. Awwad, and A. A. Mohammed, “A Robust Method to Detect Hidden Data from Digital Images,” vol. 2012, no. April, pp. 91–95, 2012.

[19] \P. Bharti and R. Soni, “A New Approach of Data Hiding in Images using Cryptography and Steganography,” vol. 58, no. 18, pp. 1–5, 2012.

[20] A. Varna, S. Rane, and A. Vetro, “Data Hiding In Hard-Copy Text Documents Robust to Print , Scan and Photocopy Operations,” 2009.

[21] A. Mohanachandran and M. L. P. A, “Secure Text / Image Data Hiding in Images with Efficient Key Management,” vol. 4, no. 5, pp. 2325–2330, 2015.

[22] A. T. Abbasi and B. Ahmad, “Urdu text steganography : Utilizing isolated letters,” vol. 2015, pp. 37–46, 2015.

[23] R. Kumar, S. Chand, and S. Singh, *“An Email based high capacity text steganography scheme using combinatorial compression,”* Proc. 5th Int. Conf. Conflu. 2014 Next Gener. Inf. Technol. Summit, pp. 336–339, 2014.

[24] K. F. Rafat and M. Sher, “StegRithm : Steganographic Algorithm for Digital ASCII Text Documents,” vol. 4, no. 6, pp. 765–769, 2012.

[25] S. Sreekumar and V. Salam, “Advanced Reversible Data Hiding With Encrypted Data,” vol. 13, no. 7, pp. 310–313, 2014.

[26] S. Gupta and R. Jain, *“An innovative method of Text Steganography,”* Proc. 2015 3rd Int. Conf. Image Inf. Process. ICIIP 2015, pp. 60–64, 2016.

[27] M. P. Uddin, M. Saha, S. J. Ferdousi, M. I. Afjal, and M. A. Marjan, “Developing an efficient solution to information hiding through text steganography along with cryptography,” 2014 9th Int. Forum Strateg. Technol. IFOST 2014, pp. 14–17, 2014.

[28] K. L. Kermanidis, “Hiding Secret Information By Automatically Paraphrasing Modern Greek Text With Minimal Resources,” 2010.

[29] C. Lin, L. Yang, and Y. Chen, “Data Hiding Scheme based on Features of Chinese Text,” pp. 0–3, 2011.

[30] [T. Topi, “Utilization of Maximum Data Hiding Capacity in Object-based Text Document Authentication,” pp. 5–8, 2006.

[31] C. Jin, D. Zhang, and M. Pan, “Chinese Text Information Hiding Based on Paraphrasing Technology,” pp. 39–42, 2010.

[32] D. Changyan, “A Data Hiding System Based on Length of English Text,” pp. 161–166, 2010.

[33] C. Science, A. K. Sahu, and G. Swain, “A Review on LSB Substitution and PVD Based Image Steganography Techniques,” vol. 2, no. 3, pp. 712–719, 2016.

[34] E. Alrashed and S. S. Alroomi, “Hungarian-Puzzled Text with Dynamic Quadratic Embedding Steganography,” vol. 7, no. 2, pp. 799–809, 2017.

[35] [F. Akhter, “A Secured Word by Word Graph Steganography using Huffman Encoding,” pp. 9–12, 2016.

[36] F. Huang, J. Huang, and Y. Shi, “New Framework for Reversible Data Hiding in Encrypted Domain,” vol. 11, no. 12, pp. 2777–2789, 2016.