Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 1-7

DOI: https://doi.org/10.25195/ijci.v50i2.487
Print ISSN: 2313-190X, Online ISSN: 2520-4912

Research Article

Maximum Error Insertion (MEI): A Novel Benchmarking Method for Data Hiding Algorithms

Baydaa Mohammad Mushgil
Information and Communication Dept.
Al-khawarizmi College of Engineering, University of Baghdad
Baghdad, Iraq
baydaait@kecbu.uobaghdad.edu.iq

ARTICLEINFO
Article History
Received: 07/04/2024
Accepted: 17/05/2024
Published: 27/5/2024

ABSTRACT

Data hiding is becoming increasingly important due to the growing threat to the privacy and security of data from intruders and hackers. This situation is accompanied by the advancement in artificial intelligence applications designed to reveal hidden data, making it difficult to choose the most appropriate hiding approach from those presented in literature. The benchmarking method serves as an important roadmap for making decisions. We propose a distinct plain benchmarking method called maximum error insertion (MEI) benchmarking. This approach intends to hide data using maximum error insertion. The MEI refers to the maximum amount of distortion that can be added to host data (such as image or audio) while still ensuring the successful retrieval of hidden data. The maximum error that can be generated by each hiding algorithm is intentionally inserted to the media file, thus giving us maximum error, maximum capacity, and maximum sensitivity to signal processing attacks. Investigation of the two hiding algorithms demonstrates their applicability and precision, and their implementation significantly enhances the reliability of results during the benchmarking stage.

Keywords:* *benchmarking; maximum error insertion (MEI); data hiding.

1. INTRODUCTION

Data hiding is a broad branch of science that primarily aims to use signal processing operations or manipulations to enable any signal to carry imperceptible information data. Depending on the purpose of data hiding, any hiding algorithm must meet one criterion or more to efficiently achieve that purpose [1-5]. The efficiency of data hiding algorithm is determined by a few specifications, namely: imperceptibility, security, robustness, capacity, and complexity [6-9]. Figure 1 illustrates the correlation between the efficiency of any algorithm and the hiding criteria. These criteria can be explained as follows: Imperceptibility refers to the quality of the cover file; after hiding data, the resultant file should appear exactly like the original one without obvious distortion [10, 11]. Robustness is the resilience power of an algorithm that allows hidden data to withstand the attacks without disappearing [12, 13]. Capacity is the size of hidden data that is measured depending on the type of cover media file [14-16]. Complexity is the amount of necessary resources that is needed to create a hiding algorithm. Security of hidden data is the impossibility of revealing data without permission [9]. However, complexity and security are often secondary criteria in the efficiency evaluation [17, 18]. The relationship between the data hiding criteria is a compromising one. Researchers are continually investigating approaches in the literature to achieve the best trade-off between these criteria. Several papers in the

DOI: https://doi.org/10.25195/ijci.v50i2.487
Print ISSN: 2313-190X, Online ISSN: 2520-4912

literature discussed data hiding algorithms, claiming to achieve better results than others [19-21]. However, no standard benchmarking platform can justify the comparisons among offered schemes.

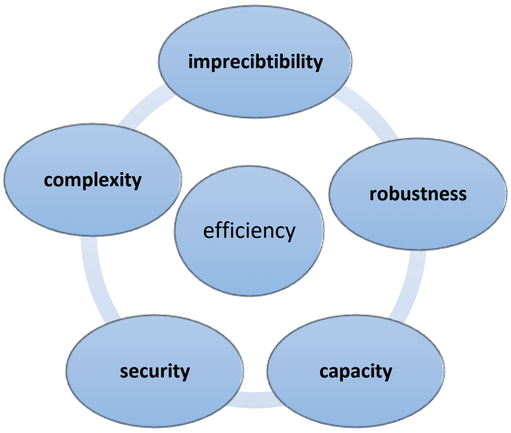

Fig. 1. Data Hiding Efficiency Criteria

2. LITRATURE REVIEW

Several benchmarking approaches are suggested in the literature. Reference [22] proposed OR-Benchmark, which is a reconfigurable benchmarking framework for most digital watermarking purposes. This framework can be considered a valuable benchmarking platform, as the authors claim it supports all benchmarking goals. Two main features are considered: first, the interfaces of the platform are public and easy to reach; second, different authors may expand the implementation and re-configure hiding algorithms. However, uncovering any hiding algorithm on a public platform is not recommended.

Reference [23] proposed a multi-dimensional matrix evaluation for benchmarking the algorithms of data hiding. However, numerous techniques of investigation are used without establishing a general one that can be adopted as the proper benchmarking.

Reference [24] defined a new metric named “Combined Capacity–Quality–Robustness Effectiveness (CCQRE)”, which integrates the measures of the three mentioned criteria into a single metric criterion. This approach neglects the existence of applications that need only one or two major criteria.

Reference [25] proposed a method for standardizing the evaluation of watermark robustness. The aforementioned study argued an important viewpoint that many other benchmarking systems lack evaluation of: the ability to detect the watermark and the necessity of maintaining a trade-off between the reliability of the detection process of the watermark and the practical advantages of the image. Nevertheless, the proposed method focused on image watermarking with only two purposes.

Despite the frequent attempts to suggest a standard benchmarking approach, many of the proposed algorithms still lack the perfection required to become a standard. Performing fair benchmarking requires implementing algorithms with all the relevant special details. Implementation of algorithms from the literature for benchmarking is difficult and necessitates a substantial amount of time. Accordingly, this work aims to propose an achievable benchmark method that can be quickly implemented, easily used by future proposed data hiding algorithms, and benchmark older algorithms. This approach aims to enhance the credibility of the benchmarking results.

3. METHODOLOGY

Benchmarking systems require implementing the data hiding algorithm and computing the benchmarking metrics. The proposed MEI-benchmarking approach imposes a slight change to the hiding process. The data hiding process must include the step of computing the maximum error to hide the data that will cause this error. A number of general metrics should be calculated to study the benchmark. After benchmarking, the user should be able to choose one of several algorithms based on the application requirements.

DOI: https://doi.org/10.25195/ijci.v50i2.487
Print ISSN: 2313-190X, Online ISSN: 2520-4912

3.1. MEI Generalization

First, the cover media files should be standardized "unified" (i.e., for image data hiding, a chapter of benchmark results must include some general cover images that are previously determined in the benchmarking platform). Most algorithms hide different data to show results or to benchmark. The precision of the results can be directly affected by the sequence of hidden bits, particularly in terms of robustness and invisibility. This situation is due to the neglect of the coincidental similarity between cover and hidden data. A new benchmarking algorithm is proposed, which is easy to follow and can be applied to all data hiding algorithms. The idea is summarized by hiding the data that cause maximum error to the cover file in the hiding process. The inserted error has the following important features: the maximum capacity, the maximum distortion caused by the hiding process (quality measures), and the maximum sensitivity to attacks (robustness measures). Figure 2 shows the general MEI benchmarking steps.

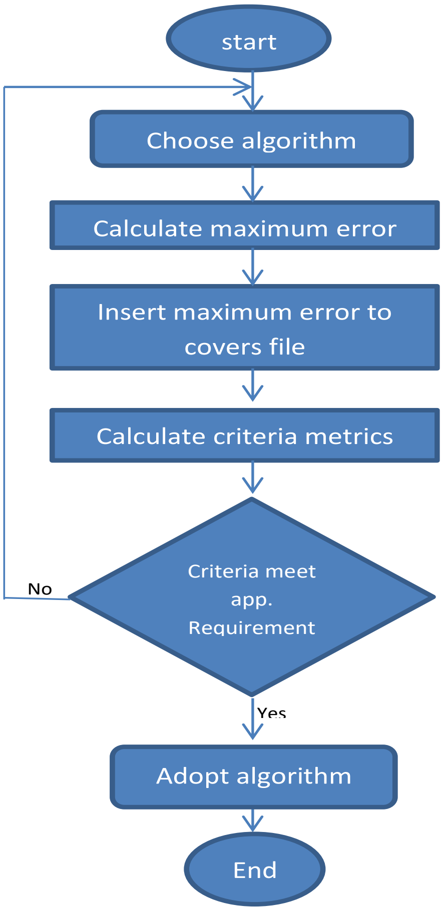
start
Choose algorithm
Calculate maximum error
Insert maximum error to covers file
Calculate criteria metrics
Criteria meet app. Requirement
No
Yes
Adopt algorithm
End

Fig. 2. General MEI Benchmarking Steps

The investigation of data hiding benchmarking algorithms starts benchmarking process with the famous least significant bit algorithms. The two famous and widely used images in the literature, namely, Lena and peppers images, are selected as cover files for benchmarking image data hiding algorithms. Figures 3 and 4 show the Lena and pepper images used. The use of these same images makes it easier to compare with previous work that investigated these images. Each image used in the investigation is of size (512*512) pixels.

3.2. Benchmarking Metrics

A substantial amount of studies in the literature focus on some metrics depending on the purpose of data hiding. The general investigation metrics in this work depend on the importance of these metrics in literature [26-29].

Fig 3. Lena Image

Fig 4. Pepper Image

3.2.1 Imperceptibility

Imperceptibility is presented by several important factors:

- **Mean squared error (MSE):** it is the average of the difference between original and manipulated image pixels. MSE metrics are typically defined as objective measures of distortion due to the simple calculations. An image with a higher MSE will show more sensible or visible distortion than the one with a lower MSE. A higher value of MSE means a greater difference between the original image and the changed image.
- Signal-to-noise ratio (SNR): calculates the strength of the image imperceptibility relative to the noise caused by the hiding process.
- **Peak SNR (PSNR):** In image and video peak SNR (PSNR) are commonly used in practice rather than MSE to characterize the reconstructed image quality. This situation is due to the PSNR that will normalize MSE the with reference to the value of peak signal instead of the signal variance. Moreover, this study enables straight comparisons among the results using diverse codecs or systems.
- **Structural similarity of image (SSIM):** a metric based on perception to consider the degradation of an image as sensed difference in the structure of the information while also comprising important imperceptibility criterion, including luminance-masking and contrast-masking terms. The importance of this technique comes from the difference with other techniques. MSE and PSNR are imperceptibility metrics that are used to estimate absolute errors. Structural information is the idea that a pixel has strong dependencies with other pixels, especially with those who are locative close. The pixel's dependencies carry crucial information about the objects' structure in the visual image. Luminance-masking is a phenomenon where distortions in an image become more imperceptible in the brighter regions. Meanwhile, contrast-masking occurs when distortions in an image become more imperceptible in areas with significant "texture" activity.

3.2.2. Capacity of Hidden Data

Capacity refers to size of the data that can be hidden in an image by using the hiding algorithm. This factor is calculated in bits per pixel (bpp).

3.3. LSB Algorithms Using MEI Benchmarking

The least significant bit (LSB) is the first algorithm to be investigated. We start with time domain LSB algorithm and frequency domain to view benchmarking figures. The LSB algorithm is one of the oldest proposed approaches in the data hiding algorithms. However, this algorithm is still being used to date [30,32]. The LSB algorithm is summarized in data hiding by substituting the least significant bit of the value of a signal with a hidden bit (i.e., in RGB images). Three LSBs of each pixel can hold three hidden bits. Consequently, the capacity of LSB is triple the size of an image. However, the time domain LSB method is not robust and can be used for fragile applications only. The frequency domain adds more robustness to the hidden data. Consequently, this work investigates the imperceptibility and capacity of the LSB method in time and frequency domain.

4. RESULTS FOR MEI-LSB ALGORITHM BENCHMARKING

The MEI in the LSB algorithm can be obtained by reversing all the LSBs of the corresponding hiding plane. Applying the LSB reversing algorithm in the time domain is easier than in the frequency domain. However, in the time domain, the MEI can be achieved directly by reversing the LSB bits as all pixels are represented as integer values. Meanwhile, in the frequency domain, the MEI will need more steps, similar to the hiding steps. The capacity obtained in the frequency domain is lower than that in the time domain as a result of reducing the size of hiding plane. Moreover, the additional steps in the frequency domain (as we use discrete wavelet transform [DWT]) will generate additional distortions to the cover image. Table 1 shows the imperceptibility metrics for the Lena and Pepper images. Meanwhile, Table 2 shows the capacity for the Lena and Pepper images.

5. IMPERCEPTIBILITY METRICS FOR MEI-BENCHMARKING.

Image	/. MEI-LSB-domain	. SSIM	. MSE	0. SNR	11. PSNR
Lena	Time	0.9996	1	6.81e−04	0
Lena	Frequency (DWT)	1	31.0562	5.12e−04	−14.921
Pepper	Time	0.9962	1	0.0032	0
Pepper	Frequency (DWT)	1	16.2639	0.0124	−12.112

TABLE I. CAPACITY BENCHMARKING FOR MEI-BENCHMARKING

Image	MEI-LSB-domain	Capacity
Lena	Time	512*512*3
Lena	Frequency (DWT)	256*256*3
Pepper	Time	512*512*3
Pepper	Frequency (DWT)	256*256*3

A decision can be made depending on the benchmarking results for applications that require a trade-off between the efficiency criteria. The results show the maximum expected values of distortion generated by both algorithms. The results in Tables 1 and 2 demonstrate that the capacity metric in the time domain is better than that in the frequency domain. Meanwhile, the imperceptibility improves in the frequency domain. This phenomenon is related to the distribution of noise all over the frequencies of the signal. The possibility of altering data by users will not give different or worse results with the studied metrics due to the use of maximum error insertion with the highest capacity available.

Although the other benchmarking attempts show only the successful side of the acquired results, the proposed MEI benchmarking gives improved method to evaluate efficiency metrics compared with these attempts, because any change in the hidden data will not change the measured MEI-efficiency metrics of the algorithm. Consequently, the proposed MEI benchmarking outperforms exiting methods stabilizing the measured metrics for any hiding algorithm. The MEI benchmarking is more effective than the existing methods because the authors will no longer need to implement other algorithms for benchmarking to unify the compared element variables. Presenting MEI-results straightway for each new introduced hiding algorithm makes it easy to benchmark any other algorithm used MEI as metric results.

DOI: https://doi.org/10.25195/ijci.v50i2.487
Print ISSN: 2313-190X, Online ISSN: 2520-4912

6. CONCLUSIONS

This work introduced the MEI benchmarking algorithm as a new platform for measuring data hiding metrics. This algorithm allows comparing various data hiding schemes even if they are working on different ranges of values for their criteria. The proposed MEI benchmarking overcomes the lack of similarity-fairness in the benchmarking. The implementation of the MEI algorithm has demonstrated the ease and credibility of the proposed approach as a platform tool for benchmarking data hiding schemes. The advantage of MEI benchmarking because of its immunity to the alterations of the hidden data. The hiding algorithm's performance will not be affected by the changing of type or size of the hidden data. Consequently, the suggested MEI benchmarking performs better than the other strategies as it stabilizes the measured metrics. In future work, the detailed robustness study may be accomplished to set a general comparison added to the imperceptibility and capacity comparison platforms.

References

[1] B. Lin and A. Li, "Study on Benchmark System for Copyright Marking Algorithms of GIS Vector Data." In 2010 18th IEEE International Conference on Geoinformatics 2010, pp.1-5

[2] T. Tuncer and E. Avcı, "A new data hiding algorithm based on direction vector for digital images," in 2015 23nd Signal Processing and Communications Applications Conference (SIU), 2015, pp. 1753–1756. doi: 10.1109/SIU.2015.7130192.

[3] S. Singh, A. K. Singh, and S. P. Ghrera, "A recent survey on data hiding techniques," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 882–886. doi: 10.1109/I-SMAC.2017.8058306.

[4] B. M. Mushgil, W. A. W. Adnan, S. A.-R. Al-Hadad, and S. M. S. Ahmad, "An efficient selective method for audio watermarking against de-synchronization attacks," Journal of Electrical Engineering and Technology, vol. 13, no. 1, pp. 476–484, 2018.

[5] M. Begum and M. S. Uddin, "Digital Image Watermarking Techniques: A Review," Information, vol. 11, no. 2, 2020, doi: 10.3390/info11020110.

[6] A. K. Singh, M. Dave, and A. Mohan, "Wavelet Based Image Watermarking: Futuristic Concepts in Information Security," Proceedings of the National Academy of Sciences India Section A - Physical Sciences, vol. 84, no. 3. National Academy of Sciences India, pp. 345–359, Sep. 01, 2014. doi: 10.1007/s40010-014-0140-x.

[7] S. K. Moon and R. D. Raut, "Information security model using data embedding technique for enhancing perceptibility and robustness Information security model using data embedding technique," 2019.

[8] S. Wadhera, D. Kamra, A. Rajpal, A. Jain, and V. Jain, "A Comprehensive Review on Digital Image Watermarking," *arXiv preprint arXiv:2207.06909*, 2021

[9] Z. Wang et al., "Data Hiding with Deep Learning: A Survey Unifying Digital Watermarking and Steganography," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.09287

[10] T. T. Takore, P. Rajesh Kumar, and G. Lavanya Devi, "A new robust and imperceptible image watermarking scheme based on hybrid transform and PSO," International Journal of Intelligent Systems and Applications, vol. 10, no. 11, pp. 50–63, Nov. 2018, doi: 10.5815/ijisa.2018.11.06.

[11] X. Zhou, Y. Ma, Q. Zhang, M. A. Mohammed, and R. Damaševičius, "A Reversible Watermarking System for Medical Color Images: Balancing Capacity, Imperceptibility, and Robustness," Electronics (Basel), vol. 10, no. 9, 2021, doi: 10.3390/electronics10091024.

[12] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images," in 1996 IEEE Digital Signal Processing Workshop Proceedings, 1996, pp. 37–40. doi: 10.1109/DSPWS.1996.555454.

[13] R. Kumar and K.-H. Jung, "Robust reversible data hiding scheme based on two-layer embedding strategy," Inf Sci (N Y), vol. 512, pp. 96–107, 2020.

[14] L. Cao, C. Men, and R. Ji, "High-capacity reversible watermarking scheme of 2D-vector data," Signal Image Video Process, vol. 9, no. 6, pp. 1387–1394, Sep. 2015, doi: 10.1007/s11760-013-0606-3.

[15] C.-C. Lin and P.-F. Shiu, "HighCapacity Data Hiding Scheme for DCT-based Images.," J. Inf. Hiding Multim. Signal Process., vol. 1, no. 3, pp. 220–240, 2010.

[16] D. Tong, C. Zhu, N. Ren, and W. Shi, "High-capacity and robust watermarking scheme for small-scale vector data," KSII Transactions on Internet and Information Systems, vol. 13, no. 12, pp. 6190–6213, Dec. 2019, doi: 10.3837/tiis.2019.12.022.

[17] J.-M. Guo and J.-J. Tsai, “Reversible data hiding in low complexity and high quality compression scheme,” Digit Signal Process, vol. 22, no. 5, pp. 776–785, 2012, doi: https://doi.org/10.1016/j.dsp.2012.04.004.

[18] C. F. Lee and H. L. Chen, “A novel data hiding scheme based on modulus function,” Journal of Systems and Software, vol. 83, no. 5, pp. 832–843, May 2010, doi: 10.1016/j.jss.2009.12.018.

[19] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, “Reversible data hiding,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 3, pp. 354–361, Mar. 2006, doi: 10.1109/TCSVT.2006.869964.

[20] W. Zhang, K. Ma, and N. Yu, “Reversibility improved data hiding in encrypted images,” Signal Processing, vol. 94, no. 1, pp. 118–127, 2014, doi: 10.1016/j.sigpro.2013.06.023.

[21] P. C. Mandal, I. Mukherjee, and B. N. Chatterji, “High capacity reversible and secured data hiding in images using interpolation and difference expansion technique,” Multimed Tools Appl, vol. 80, no. 3, pp. 3623–3644, 2021.

[22] H. Wang, A. T. Ho, and S. Li, “OR-Benchmark: An Open and Reconfigurable Digital Watermarking Benchmarking Framework,” May 2015, [Online]. Available: http://arxiv.org/abs/1506.00243

[23] B. B. Zaidan, A. A. Zaidan, H. Abdul Karim, and N. N. Ahmad, “A New Approach based on Multi-Dimensional Evaluation and Benchmarking for Data Hiding Techniques,” Int J Inf Technol Decis Mak, pp. 1–42, 2017, doi: 10.1142/S0219622017500183.

[24] T. Rabie, M. Baziyad, T. Bonny, and R. Fareh, “Toward a unified performance metric for benchmarking steganography systems,” Journal of Circuits, Systems and Computers, vol. 29, no. 03, p. 2050042, 2020

[25] B. An et al., “Benchmarking the Robustness of Image Watermarks,” arXiv preprint arXiv:2401.08573, 2024.

I. F. Ahmed, “Hide text in the WAV audio file,” Journal of Education and Science, vol. 24, no. 4, pp. 154–168, 1970.

[27] A. K. Hmood, H. A. Jalab, Z. M. Kasirun, A. A. Zaidan, and B. B. Zaidan, “On the capacity and security of steganography approaches: An overview,” Journal of Applied Sciences, vol. 10, no. 16, pp. 1825–1833, 2010.

[28] A. N. Netravali, Digital pictures: representation, compression, and standards. Springer, 2013.

[29] D. R. Bull and F. Zhang, Digital picture formats and representations. Communicating pictures, 99-132, 2014.

[30] M. D. Giakoumakis, “Refinements in a DCT based non-uniform embedding watermarking scheme,” Monterey, California. Naval Postgraduate School, 1982.

[31] S. A. Jebur, A. K. Nawar, L. E. Kadhim, and M. M. Jahefer, “Hiding Information in Digital Images Using LSB Steganography Technique.,” International Journal of Interactive Mobile Technologies, vol. 17, no. 7, 2023.

[32] A. K. Saini and S. Singh, “HSB based reversible data hiding using sorting and pairwise expansion,” Journal of Information Security and Applications, vol. 80, p. 103663, 2024.