Journal of Iraqi Al-Khwarizmi (JIKh) Volume:6 Issue:2 Year: 2022 pages: 37-43

Review: Secure data hiding by machine learning in cloud

[1]Hiba Hamdi Hassan, [2]Maisa'a Abid Ali Khodher

[1,2]Department of Computer Sciences
University of Technology –Iraq
Baghdad/Iraq
[1]Hiba.albairam@gmail.com [2]110044@uotechnology.edu.iq

Published online 28-11-2022

Abstract: Significant quantities of data are generated by systems focused on fog computing; An growing number of applications and resources for cloud are thus emerging. In addition, machine learning (ML), an important field, has made considerable advances in various fields of research such as computer graphics, decision-making, and several studies have proposed researching how to use ML to resolve issues with cloud. The cloud concerns, when people use the internet (internet computing) to increase in the present time, must also secure data during transfer from sender to receiver. And in this article, In solving this issue in the cloud, multiple techniques are used. This paper, by using unsupervised cluster machine learning techniques to conceal hidden message in Images and store secret data in cloud.
Keywords: *Cloud, Machine Learning, Steganography, Clustering, Data hiding.*

1. INTRODUCTION

Data hiding methods make it possible to encode hidden messages into Objects of an Undetectable way, so the main image and the identified image (embedded data image) appear physically identical. The discrepancy between the the immediate image and the categorized image is known as noise, so any third party cannot quickly recover the encoded message. In current data hiding methods, the number of bits in an image, however, also known as the incrustation power, is much limited. Built in message is therefore very short (only few characters) [1]. Steganography uses a hidden key for the method of decryption. The message can't be decrypted without the decryption key. Any of the examples are LSB algorithms, MSB algorithms, Huffman code algorithms, and several algorithms used for image steganography [2]. Although steganography is designed to do embedding Difficult to know, steganalysis aims to detect, not actually recover, the presence of a secret message. Steganalysis depends on the plain fact that any distortion can occur in every embedding scheme. There are two basic methods, namely systemic and mathematical, for stegnalysis using local pixel similarity, structural techniques detect changes, and statistical techniques examine the Objects of statistics of an image to detect differences in the statistical property of the stego image [3]. Steganography primarily attempts to preserve the system's robustness and thus ensure that the representation of the stego is less visible to the eyes of people. It will be advantageous to mask the hidden material with less distortions to the cover image with good payload capability and fewer user end recovery error by choosing the coefficient using machine learning techniques. The stego picture renders these elements [4]. Cloud services are becoming irreplaceable, which has added to the importance of connectivity and data protection. Either of two approaches or variations will accomplish data encryption for cloud storage, Cryptography is the first, and the other secret data. In line with different implementations [5].

2. THE LITERATURE VIWER

The work of previous writers where text and image steganography, machine learning in cloude, and clustering K- means algorithm.

In 2016. Khalid A. Al-Afand, and et.al. Proposed using the cropping picture and Least Significant Bit (LSB) steganography, a highly secure data hiding technique. It is focused on splitting the hidden text message into four sections and extracting, with some secret coordinates, four crops from the cover color graphic. The result is a more reliable data hiding approach and a more complicated secure data retrieval method at the same time. [6].

In 2018. Amjad Rehman CCIS, and et.al. The proposed algorithm used the decomposition of the Fibonacci number, which increased the cover picture of the bit planes. As a result, our methodology has the potential to create a high-quality stego image and have greater protection relative to other techniques. The outcome of the suggested approach is simple and reliable and possible for the application of steganography. [7].

In 2018.Wang GAO, Min Peng1 and et .al. proposed Design a novel model, known as Conditional Random Field regularized Topic Model, for short text topic modeling (CRFTM). Not only does CRFTM create a simplified approach by aggregating short texts into pseudo-documents to mitigate the sparsity problem, but it also uses a regularized Conditional Random Field model that allows semantically similar terms to share the same subject assignments [8].

In 2019.Florent Poux and Roland Billen Proposed a function engineering based on voxel that better characterizes point clusters and offers supervised or unsupervised classification with good support. In order to enable interoperable systems, we have numerous feature generalization levels. Second, a form-based function collection (SF1) which leverages only raw X, Y, Z attributes from any point of the cloud. Product of A classification method to automatically detect the principal classes in the S3DIS dataset and to achieve a performance measure against best-performing deep learning method [9].

In 2019.Shipra Varshney proposed the techniques On the basis of the data collection, the sensitive item is increased and optimally hidden, also in large posts, which ensures optimum hiding time. The results of the discovered techniques showed that techniques were executed in order to understand the accuracy of privacy, the time needed for optimum data hiding and the degree of side effects of changed data sets compared to the state of the art works [10].

in2019. Ahmed A. Abd El-Latif ,Bassem Abd-El-Atty and et .al . Proposed the main concept for the presentation of these algorithms is to integrate the sensitive object into the host media without the hidden data pre-encryption process, which is entirely focused on quantum walks in its protection and embedding processes. It also opens the door for quantum technology to be paired with information hiding strategies to achieve better protection. [11].

in 2019. Nico Verbeeck , Richard M. Caprioli and et.al Proposed the In order to obtain useful information from the broad and high-dimensional datasets obtained from IMs, unsupervised data analysis approaches have become essential.. In order to obtain useful information from the broad and high-dimensional datasets obtained from IMs, unsupervised data analysis approaches have become essential [12].

in 2020. Jingli Ren Proposed a systematic analysis of data protection and privacy problems literature, Technology for Encrypting Data, and relevant cloud storage system countermeasures. In particular, we first give A Cloud Computing Overview, including description, Technology and software. Secondly, we offer a thorough overview of data management and confidences protection issues and specifications in the cloud storage environment. Thirdly, tools for data encryption and strategies of security are summarized. Finally, some open research subjects on data protection for cloud computing are discussed. Result of Technologies for data encryption and methods of security are resumed. These conform to the safety criteria Specified. Addressed a range of open data security analysis themes for cloud computing [13].

3.STEGANOGRAPHY

Steganography is a technology that covers hidden data in clandestine transmission coverage. Steganography's most critical goal is to counteract the identification of the enemy using steganalysis techniques. Many of the strategies that function for photographs are more nuanced when applied to natural language text as a cover tool, with an emphasis on hiding data in images many steganalysis methods aim to find statistical irregularities in the cover data that forecast the existence of secret knowledge [14, 15].

3.1. IMAGE STEGANOGRAPHY

Image steganography has been more popular with the press than other forms of steganography, probably due to the influx of Electronic image data available with digital and high-speed internet cameras streaming. Image steganography also entails covering details inside the image in the "noise" that naturally occurs and offers a clear example for those techniques [14]

3. 2. TEXT STEGANOGRAPHY

Text is still one of the oldest media used in steganography; letters, books, and telegrams concealed hidden messages within their documents long before the electronic era. In this part, we will explore more in depth text steganography, address the state-of-the-art and incorporate current linguistic methods. This approach produces cover text by random sequences of characters, altering words in text, using context-free grammars, or Optimizing current document formatting to mask letters. The text of the cover developed by this method will Language steganography certification when the text is directed linguistically. While this text-based approach has its own distinctive cover text features, from both a linguistic and a security viewpoint [14].

4. CHARACTERIZATION OF STEGANOGRAPHY SYSTEMS

4.1. CAPACITY

The idea of data hiding capacity implies the cumulative amount of bits shielded and retrieved effectively by the Stego device [15].

4.2. ROBUSTNESS

refers to embedded data's capacity to stay Static while the stego system is transformed, such Linear and nonlinear filtration; random noise addition; and compression scaling, rotation, and loose [15].

4.3. THE EMBEDDED ALGORITHM

if the embedded message image is compatible with the source model the images are taken from, it is undetectable. For instance, if a steganography system utilizes the noise portion of digital images to embed a hidden message, it may do so without requiring statistical improvements to the noise of the carrier. The size of the secret message and the format of the cover image content is directly affected by the potential of the undetected [15].

5. SPATIAL TECHNIQUE:

5.1 . LSB TECHNIQUE

Embedding can be accomplished by simply substituting the hidden message for the unknown LSB chosen pixel in the cover graphic bit. Variety of steganography based on LSB. Some of them have an adaptive LSB substitution to estimate the numbers k for data sheds based on the host image's brightness, borders and texture masking. suggested approach is based on the idea that a smaller number of changes than heavily textured areas can be accepted by edge areas and not more changes than smooth areas. n case, the method integrates more hidden data into non-sensitive noise areas than sensitive noise areas. Multi-bit plane steganography: This approach provides an extension of the basic technique of LSB substitution. In multiple-bit aircraft, coded message bits are concealed [16, 17]

5.2. GRAY LEVEL

This approach is used for mapping data by modifying the gray pixel levels (not adding or hiding them). Gray level resolution refers to the smallest variation in the gray shades or degrees image. a group of pixels is chosen for mapping depending on some mathematical function. The definition of odd and even numbers is used by this technique to map data inside an image. The darkest color is black, and white is the lightest gray color range. The classification of the gray scale is more difficult than human eye color detection. Low computing complexity and high data hiding capability are advantages of this approach. [16, 18].

6. STEGANOGRAPHY TRANSFORM DOMAIN

6.1. THE INTEGER WAVELET TRANSFORM (IWT)

Is a process that maps the binary data set with a different integer date set. The main property of IWT is that coefficients are dynamic as the primary signals. The number of variables it uses and the fields it gives are taken into consideration in the code algorithm for speedy execution. Included are four convergent, longitudinal, horizontal and diagonal stripes that strongly appear as LL, LH, HL and HHH. [19].

6.2. DISCRETE WAVELET TRANSFORM (DWT)

Sampling of wavelets by any wavelet transform is DWT this is favored over Fourier Transforms because temporal resolution is the biggest benefit. Both frequency and position are recorded by DWT. It allows time and frequency signals to be processed simultaneously. In comparison to DCT, DWT forms the basis for JPEG2000 image compression [20]. the wavelet transform was used for the frequency domain. The use of wavelets in the simplified measuring model depends on the original statement that the wavelet process of transition distinguishes the higher from the less reserving pixel-based information. [19]. Results in four subsets of image reducing. The lower subset has the most necessary data, and the higher subset has better information. most energy is converted into a few coefficients, found and encrypted by a code of entropy. DWT offers better energy compaction compared with DCT, and without blocking the artifact after coding, DWT breaks the picture down like the dyadic pyramid at L-level. The wavelet coefficient can be easily scaled out in resolution as the wavelet coefficients can be removed to a fixed value at levels that are thinner and the image reconstruction can therefore be done with less detail. [16].

6.3. THE (DCT) DISCRETE COSINE TRANSFORM

JPEG is significant today and the file format is widely used on the internet. JPEG uses DCT spatially to transform the domain. DCT takes correlated data and concentrates its energy in the first few transformation coefficients. Image compression is a pixel association based on two dimensions a pixel does not seem to just mimic those in their lines but is all similar to their neighbor.2D DCT is often used. An foreign image compression form within the Joint photographic expert group minimized the blocking effect of image compression (JPEG). The two-dimensional DCTs were decomposed into 1 pair of single (1D) CTs in the algorithm during transformation for a JPEG image of (8*8) the volume of space blocks. 2D spatial data is a linear mix of the law image produced by the outer effects and vectors of the cosine function column so that the opposite DCT is as active as the reverse DCT [19]. The Separate Convolution Transformation (DCT) divides the image in parts or spectral sub bands of different value with respect to its optical consistency. The DCT is related to Fourier's discrete transformation, which transforms a spatial domain signal or image into a frequency. [18].

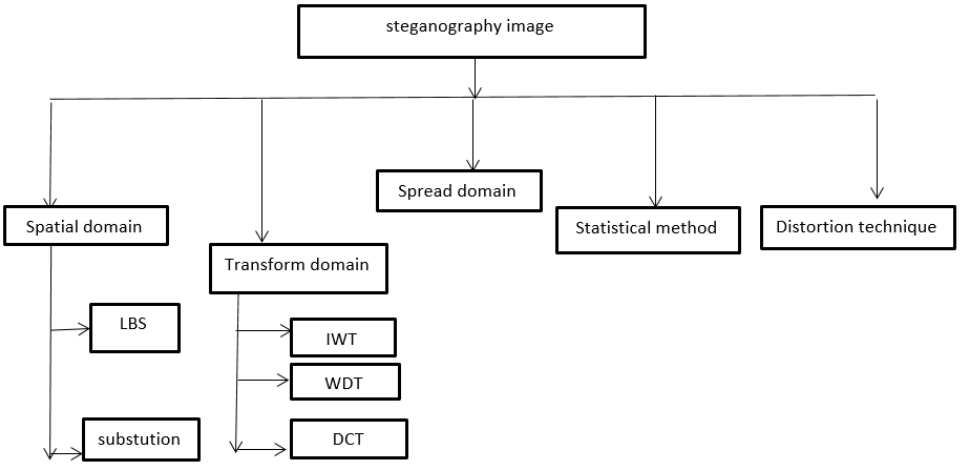

FIGURE 1: The steganography image methods.

7. MACHINE LEARNING

It's about developing algorithms that derive useful knowledge from data automatically. The focus here is on "automatic," i.e., machine learning is concerned with methodologies of general intent that can be extended to multiple datasets, while generating something significant. The heart of machine learning is comprised of three concepts: data, a model [21].

7.1. SUPERVISED LEARNING

The goal of the algorithms is to construct a model to map output information, which is called "monitored learning" by completely labeling the attributes of the input and output data sets. In supervised learning, classification, regression, etc. are representational applications. Two conventional control procedures for education [21].

7.1.1. CLASSIFICATION TECHNIQUES

Predict, for example, different responses whether an original email is actual or spamming, whether a tumor is benign or cancerous. Input data is grouped into classes by classification models. Medical imaging, speech recognition and credit rating are widely used. [21]**.**

7.1.2. REGRESSION

For starters, techniques forecast continuous reactions, temperature changes or variations in generating capacity. redicting energy load and algorithmic exchange are common uses [21]**.**

7.2. UNSUPERVISED LEARNING

Algorithms also, on their own, derive features and patterns. Typically, the models establish a deep correlation with the aid of internalized heuristics as per the Data input identity or distance. Clustering methods are an important algorithm in unsupervised learning [21]

7.2.1. CLUSTERING

he most popular unsupervised learning methodology is. In order to identify hidden patterns or groupings in data, it is used for exploratory data analysis. Clustering uses include gene sequence analysis, market study, and detection of artifacts [18]. Any plan or scale is perfect data analyst. But the compilation of algorithms often relies on the size and form of data, in order to learn from these data [22].

8. CLUSTERING ALGORITHIMS

1. K-MEANS

A popular commonly algorithm used is a clustering the one first suggested by James Macqueen. K-Means is one of the simplest machine learning clustering algorithms that can be used in data training to identify classes of identical instances, items, objects, points automatically. The algorithm categorizes case into a predefined number of user-specified clusters (e.g. assume k clusters). The first significant step is to pick a set of k cases as centroids (centers of the clusters) [22].

8.2. EM CLUSTERING

Expectation Maximization(EM) clustering is the k-mean clustering variant used commonly in unregulated clustering for the density estimation of data points. In the EM clustering, the parameters are used by an EM algorithm to maximize the probability

of the data, given that the data is generated from the normal k distribution system. The algorithm knows both how naturally distributed the algorithm. [22].

8.3. OUTLIER DETECTION

It is a tool for identifying data patterns that do not adhere to planned behavior Clustering algorithms are in other words designed to identify rather than outliers clusters. Most clustering Algorithms are not available allocate all points to clusters but allow for noise artifacts. Through applying one of the clustering algorithms, outlier detection algorithms search for outliers and recover the noise collection, so the efficiency how outer algorithms are detected depends effective the Algorithm of clustering limit is [22].

8.4. FUZZY CLUSTER

For expressive prosthetic hand power, sign languages, grip recognition, human-machine interaction, etc., recognizing and classifying electromyogram (EMG) signals is important [23]. The latest EMG-based hand gesture classification research faces the difficulties of unsatisfied accuracy of classification, inadequate capacity to generalize, lack of training data and limited robustness. This paper incorporates unsupervised and supervised learning approaches to define an EMGG in order to resolve these concerns [23].

9. CLOUD

A hot subject in the IT sector has been cloud computing. If computing and storage facilities like electronics and water can be quickly accessed, it would be a breakthrough in the IT industry [1]. Data is stored in a public storage provider in cloud computing environments. The most critical part of cloud storage is data protection. In many geographic regions, cloud storage service providers will create data centers. Most providers offer save and acquire service at the file level. It is crucial how to break files into pieces and how to position these pieces to make files safer. The level of file protection from low to high is split into single-server, cross-server, cross-cabinet and cross-data center tiers. Both are divided by the location of the components stored in cloud computing [1].

The key problems facing data protection and privacy preservation in the cloud computing environment are the following:

1. Fine-grained regulation of data usage.

Cloud providers malicious will false audit return reports for honesty.

2. Attack of side channels.

Cloud providers are not malicious cooperate with the wishes of clients to erase data in the cloud entirely.

3. Privacy-preservation [24].

10. CONCLUSION

It is convenient to conceal hidden message from this paper offering steganography image, since small specifics of the image cannot be visible through the human eye. But it is very difficult for hidden steganography text because the length of scale is very small and because Text can be seen by the eye of Human in various domains, Mechanical failure identification, analyzes of medical data, fuzzy c-means, how to secrete data by clustering algorithms are frequently used. In the mechanical failure diagnosis clustering algorithm. The rolling data values bearings were used to differentiate the types of faults and positive results were achieved. When analyzing complex data, however, it provides. Cover Hidden Picture is more effective large data analysis e in shielding secret messages for cloud, robustness, and high storage and high security.

REFERENCES

[1**]** Zhang, X., Du, H. T., Chen, J. Q., Lin, Y., & Zeng, L. J. (2011, May). Ensure data security in cloud storage. In 2011 International Conference on Network Computing and Information Security (Vol. 1, pp. 284-287). IEEE.

[2] Garware, N., Shinde, P., Patel, V., Patil, N., & Patil, Y. Steganography: Data Hiding using Cover Image.

[3] Chhikara, S., & Kumar, R. (2020). MI-LFGOA: multi-island levy-flight based grasshopper optimization for spatial image steganalysis. Multimedia Tools and Applications, 79(39), 29723-29750.

[4] Chanchal, M., Malathi, P., & Kumar, G. (2020, October). A comprehensive survey on Neural Network based Image Data hiding Scheme. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 1245-1249). IEEE.

[5] Abd El-Latif, A. A., Abd-El-Atty, B., Elseuofi, S., Khalifa, H. S., Alghamdi, A. S., Polat, K., & Amin, M. (2020). Secret images transfer in cloud system based on investigating quantum walks in steganography approaches. Physica A:

[6] Arulmurugan, R., Sabarmathi, K. R., & Anandakumar, H. (2019). Classification of sentence level sentiment analysis using cloud machine learning techniques. Cluster Computing, 22(1), 1199-1209

[7] Rehman, A., Saba, T., Mahmood, T., Mehmood, Z., Shah, M., & Anjum, A. (2019). Data hiding technique in steganography for information security using number theory. Journal of Information Science, 45(6), 767-778

[8] Gao, W., Peng, M., Wang, H., Zhang, Y., Xie, Q., & Tian, G. (2019). Incorporating word embeddings into topic modeling of short text. Knowledge and Information Systems, 61(2), 1123-1145.

[9] Poux, F., & Billen, R. (2019). Voxel-based 3D point cloud semantic segmentation: unsupervised geometric and relationship featuring vs deep learning methods. ISPRS International Journal of Geo-Information, 8(5), 213.

[10] Varshney, S. Hiding Techniques for Data Publishing by Conserving Item hiding and Privacy Conserving.

[11] Abd El-Latif, A. A., Abd-El-Atty, B., Elseuofi, S., Khalifa, H. S., Alghamdi, A. S., Polat, K., & Amin, M. (2020). Secret images transfer in cloud system based on investigating quantum walks in steganography approaches. Physica A: Statistical Mechanics and its Applications, 541, 123687.

[12] Verbeeck, N., Caprioli, R. M., & Van de Plas, R. (2020). Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. Mass Spectrometry Reviews, 39(3), 245-291

[13] Yang, P., Xiong, N., & Ren, J. (2020). Data security and privacy protection for cloud storage: A survey. IEEE Access, 8, 131723-131740.

[14] Bennett, K. (2004). Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text.

[15] Al-Ani, Z. K., Zaidan, A. A., Zaidan, B. B., & Alanazi, H. (2010). Overview: Main fundamentals for steganography. arXiv preprint arXiv:1003.4086.

[16] Subhedar, M. S., & Mankar, V. H. (2014). Current status and key issues in image steganography: A survey. Computer science review, 13, 95-113

[17] Yang, H., Sun, X., & Sun, G. (2009). A high-capacity image data hiding scheme using adaptive LSB substitution. Radioengineering, 18(4), 509-516.

[18] Nidhi Menon. Survey on Image Steganography. International Conference on Advancements in Power and Energy (TAP Energy), IEEE, 2017.

[19] Khairi, T. W. A. (2020, July). A comparison Steganography Between Texts and Images. In Journal of Physics: Conference Series (Vol. 1591, No. 1, p. 012024). IOP Publishing.

[20] Mayukha, S., & Sundaresan, M. (2020). Enhanced Image Compression Technique to Improve Image Quality for Mobile Applications. In Rising Threats in Expert Applications and Solutions (pp. 281-291). Springer, Singapore.

[21] Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. Information Fusion, 57, 115-129.

[22] Syarif, I., Prugel-Bennett, A., & Wills, G. (2012, April). Unsupervised clustering approach for network anomaly detection. In International conference on networked digital technologies (pp. 135-145). Springer, Berlin, Heidelberg.

[23] Xiong, J., Liu, X., Zhu, X., Zhu, H., Li, H., & Zhang, Q. (2020). Semi- Supervised Fuzzy C-Means Clustering Optimized by Simulated Annealing and Genetic Algorithm for Fault Diagnosis of Bearings. IEEE Access, 8, 181976-

[24] Yang, P., Xiong, N., & Ren, J. (2020). Data security and privacy protection for cloud storage: A survey. IEEE Access, 8, 131723-131740.