2012 International Conference on Advanced Computer Science Applications and Technologies

An Introduction to Image Steganography Techniques

Alaa A. Jabbar Altaay
1-Faculty of Information & Communication Technology
Universiti Teknikal Malaysia Melaka
2-College of Science, Al-Mustansiriyah University, Iraq
alaaj@student.utem.edu.my

Shahrin bin Sahib
Faculty of Information & Communication Technology
Universiti Teknikal Malaysia Melaka
76100 Melaka, Malaysia
shahrinsahib@utem.edu.my

Mazdak Zamani
Advanced Informatics School
Universiti Teknologi Malaysia
54100 Kuala Lumpur, Malaysia
mazdak@utm.my

***Abstract*—Steganography is a form of security technique through obscurity; the science and art of hiding the existence of a message between sender and intended recipient. Steganography has been used to hide secret messages in various types of files, including digital images, audio and video. The three most important parameters for audio steganography are imperceptibility, payload, and robustness. Different applications have different requirements of the steganography technique used. This paper intends to give an overview of image steganography, its uses and techniques.**

Keywords- digital image; information hiding; multimedia security; watermarking; steganography

I. INTRODUCTION

Art of data hiding in digital media, steganography and watermarking, aims to embed secret data into cover with purpose of identification, copyright protection, and annotation. The main constraint factors of this process are message data quantity, necessity of invariability of embedded data under distortions like lossy compression, third party removal, or modification. Data hiding techniques fall into three categories of cryptography, steganography, and watermarking. Watermarking and particularly steganography tend to conceal presence of hidden data while cryptography makes data gibberish [1, 2].

Representation of media in digital format facilitates its access and as well enhances accuracy, efficiency, and portability of existence of data. But in the opposite side undesirable effects are possibility of copyright violation and modification or tampering the content. Intellectual property protection, content manipulation indication, and annotation are the main motivations of employing these techniques [2].

Digital data hiding falls into different classes like embedding copyright information in different digital media formats such as text, audio, image, or video with the least possible perceivable degradation effects on the host signals. For example, effects must be inaudible or invisible to its observers. Data hiding techniques are distinct from encryption techniques as they aim to make the embedded data unrecoverable and inviolate-able [3, 4].

Quantity of hidden data and data invariance to manipulation requires different data embedment methodologies and as far as no single method is able to reach all of the goals then various classes of techniques are needed for spanning all ranges of applications [7].

Main usages of digital media data hiding techniques are preserving copyright and assuring content integrity. To achieve counted objectives embedded data should be kept hidden in host signal even if is subject of degrading manipulations such as lossy data compression, cropping, re-sampling, or filtering. Since embedded data are in favor of both author and consumer all data hiding applications, like augmentation data inclusion, must be invariant against removal or detection [5, 6].

Data hiding has formidable technical challenges. Perceptual or statistical holes to be filled with data in host signals are likely to be removed by means of lossy signal compression. Important factor to achieve successful data hiding technique is to find holes which are not convenient to be exploited by compression algorithms. The main challenge is filling data in this kind of holes in a way that is not easy for compression algorithms to exploit it. An enhanced challenge is filling the holes in a manner that remains invariant against signal transformation in big scale. Following counted features and restrictions are the criteria which a data embedding algorithm must meet [8]:

- Quality of host signal should not be degraded objectionably and the perceptibility of embedded data must be kept minimal.
- The data must be embedded into whole body of the target media rather than wrapper or header. Therefore it would be kept intact in different formats.
- The data must be secure against intentional and intelligent removal attempts such as filtering, encoding, cropping, channel noise, lossy compressing, re-sampling, scanning and printing, digital to analog (D/A) conversion, analog to digital (A/D) conversion, and etc.
- Since data hiding goal is to keep the embedded data into host signal, embedded data asymmetrical encoding is desirable feature but not essential.
- To guaranty data integrity error correction coding is necessary. Degradation of embedded data at signal modification time is unavoidable.
- Arbitrary re-entrant and self clocking are mandatory properties of the embedded data. These properties are to guaranty that embedded data will be retrievable even if only some fragments of the host be available.

Today there are various applications of information hiding. Knowledge of data hiding might be used either in ethical or unethical ways. However, data hiding algorithms cannot easily be categorized either in steganography or

978-0-7695-4959-0/13 $25.00 © 2013 IEEE
DOI 10.1109/ACSAT.2012.25

watermarking categories as there is no transparent boundary between these two terms and mostly the classification relies on application of the algorithm. Therefore regardless classifying data hiding the most common data hiding applications are fingerprinting, secret communication, secure storage, covert communication, and copyright protection [9].

Fingerprinting allows tracing originator or recipients of particular copy of the media be traced by means of watermarking. The employed watermarking technique must support high degree of robustness against both intentional and unintentional attacks. For instance, in advance of distributing plentiful copies of digital multimedia products among users, the copies can be watermarked by embedding various identity or serial numbers [10, 11].

Secret communication can be established by means of hiding secret information within digital media covers to hide presence of communication. This application falls in category of steganography rather than watermarking [12].

Secure storage means utilizing the cover digital media as secure storage for some sensitive information. For instance, drug prescriptions or medical records of patients need to be kept secure at storing and transmission time because of the consequences it may cause if be abused by unauthorized people for illegal activities or identity theft to fraud insurance [13].

Covert communication for some organizations or people might be vital to keep their data safe against unauthorized people. For instance, army can use this method to make covert transmission of technical information of battle plans against attackers which could harm whole operation if the data be compromised [14].

Copyright protection helps to protect dedicated resources to production of intellectual properties. Specifically in industrial societies reproduction expenditure of an intellectual property is far less than creation of it. To protect the copyright some data which represents the property owner information will be embedded within the host digital media. The watermarking algorithm is expected to be robust against attacks and let the owner prove the ownership in court cases. Fragile watermarking also might be used for host signal tampering. In addition, watermarking can be used for controlling access policy or limiting particular copies [15].

II. STEGANOGRAPHY MEASUREMENTS

As shown in Figure 1, important steganography measurements are as follow:

Figure 1. Measurement triangle of steganography

A. Capacity

Capacity is the maximum amount of secret information can be embedded in a file. Capacity either can be defined as an absolute value in term of number of bits for particular cover or as a relative number regarding necessary bits to save final stego file. Capacity value depends on both embedding function and cover properties ($x^{(0)}$). For instance, in LSB technique if the cover is 8-bit grayscale image file for one bit per pixel embedding the capacity would be equal with 12.5% or less because of the cover file header information which is not embeddable. This fact also is notable that not always the secret message is in maximum embeddable size and bit per pixel is just a measurement of capacity for maximum embedding [16]. P is a metric which shows proportion of length of the secret message relative to the maximum length of message can be embedded in cover. P value would be $0 \leq p \leq 1$ and can be calculated by following formula:

$$p = \frac{|m|}{n} \quad , \quad x^{(0)} \in x^{n}$$

Not always finding embedding rate is this easy because some stego systems are able to embed into compressed covers and therefore final embedding ratio would be variable. In these kinds of cases it is hard to find particular formula which can accurately define embedding capacity. So another capacity measurement for compressed formats is needed. For instance, F5 is a steganographic algorithm for compressed JPEG format which reduces file size monotonically with the needed size of embedding secret data [17, 18]. In fact it degrades quality of lossy compressed images to make free hole for embedding secret information.

A bit per non-zero DCT coefficient (bpc) is a capacity metric for JPEG images. Practical and theoretical studies [19] show that larger secret messages would have more changes on cover and statistically are more detectable than short ones. Therefore embedding rate and capacity are directly related to property of imperceptibility.

B. Imperceptibility

Stego object should not have important perceptual artifact. The higher fidelity of stego object, will give the better imperceptibility. This property would be satisfied if difference of resultant stego file be not distinguishable from original cover for warden [20].

There are various evaluation techniques different steganography types but the main evaluation method is PSNR. Peak Signal to Noise Ratio (PSNR) is a metric to evaluate the ratio between possible maximum signal and influence of modifying noise to fidelity of its representation [21]. This metric can be calculated as follow:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

$$PSNR = 10.\mathrm{Log}_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

C. Robustness

Robustness is property of harness of eliminating secret information from stego file. While detection of embedded secret data has much higher importance than its removal, but property of robustness talks about resisting against

intentional distortion of communication channel by means of systematic interface or channel noise aiming to ban use of steganography techniques. Robustness metrics of steganographic algorithms are classified in distortion classes like geometric transformations or additive noise. In each one of the classes distortion value can be expressed according generic (like noise source parameters) or specific (like PSNR) measures [22, 23].

Robustness of steganography methods also can be examined through steganalysis attacks. Challenging aim of steganalysis is detection of existence of the secret message in cover file [24]. Today numerous methods exist which can conduct steganalysis to reveal existence of secret information especially when the cover file is digital image [25]. However, famous steganalysis approaches are as follow [26]:

- Visual detection
- Histogram analysis (detecting according first order statistics)
- Twofold statistical techniques for images by using spatial correlations
- Higher order statistics (RS)
- Steganalysis of JPEG files' compatibility
- Universal blind detection methods

Also there is another important feature called pair of values. Chi-squared attack examines presence of PoVs to discover existence of embedded secret information.

D. Definitions in CIA triangle

In addition of capacity, robustness, and imperceptibility criteria there are more evaluation metrics which fall under each side of CIA triangle. Compression ratio, multiple watermarks, success rate, embedding complexity, and detection complexity are important definitions under integrity side [27].

Compression ratio: some of host types might be compressed like audio or image files. The host file containing embedded data must yield the same compression ratio as the unmarked one and also should not be degraded. Furthermore, the compression process should not remove the embedded data [28].

Multiple watermarks: more than one user must be capable of embedding watermark within a host file. It means that ideally a user must be capable of embedding information without changing pre-embedded one which might exist beforehand. This property must be kept even if the algorithms are not identical [29].

Success rate and embedding complexity: evaluation metrics which try to determine a scaling system for delivering desired level of security and robustness for embedment into chosen cover [10].

Detection complexity is another steganography metric which examines needed computational efforts to acquire defined mixture of (α, β) error rates.

Statistical invisibility, secret key, and secrecy are concepts and parameters involved in confidentiality side of CIA triangle. Robustness attacks are aimed to increase the value of Bit Error Ratio (BER) in watermarking or steganography channel while security attacks are designed to discover the secret key. In contrast of robustness property which is about blind attacks, concept of security is more critical factor since it talks about intentional attacks which attacker has knowledge of utilized scheme that increases likelihood of compromising [31].

Secret key: to ensure watermarking is not readable or removable the embedding algorithm must use cryptographic keys to protect it from alteration. If an attacker be able to read the embedded data it also would be able to manipulate it as both location and embedding algorithm are known. Since protection of security of embedded data is depends on the secret key then the key space must be large enough to make running brute force attacks impractical. The embedded message could be encrypted in two levels by a cipher key and data embedding key. This practice delivers two levels of security which in highest level user will not be able to decode, read, or even detect presence of embedded data. In second security layer lets any user to detect existence of embedded data, but data will not be disclosure without having the proper key [32].

Statistical invisibility: this property is needed to prevent detection of existence of steganography/watermarking. Running statistical examinations on group of watermarked files should not disclosure any information neither about employed watermarking technique nor nature of embedded data [33].

Secrecy: is defined as hardness of extracting the embedded secret message [24]. This property can be mapped to confidentiality metric of encryption systems which used beforehand of message embedment [25].

Computational cost and redundancy are two important factors under availability side of CIA triangle in field of steganography/watermarking. Computational cost: the embedment and detection time are crucial factors in data embedding algorithms. Some of applications such as broadcast monitoring need real time data processing and no delay is acceptable. In some others like court cases effectiveness is the most important factor and time could be ignored [26].

Redundancy: to enhance robustness of embedded data it could be embedded in more than one part of host file. This could happen when embedded information just takes small portion of file [27].

Steganography, watermarking, and encryption are close techniques for protecting ownership and data secrecy. Regarding the chosen technique each criterion might have different property or be called differently. For instance the secret data in steganography is called payload, in watermarking is called watermark, and in encryption is plain text. Table 2 counts important criteria of the techniques and describes property description of each [28].

III. STEGANOGRAPHY TYPES

Information hiding technology falls into three classes of steganography, watermarking, and cryptography. Steganography and watermarking each one fall into two sub-classes. Figure 2 illustrates classifications of all three classes of information hiding technology.

A. Fragile Steganography

Fragile method means insertion of information in a cover in a way that modification of host file will destroy whole embedded information. Since it can easily be removed from the career then is not convenient choice for copyright protection, but for example in court cases it can be witness of file originality. Implementation of fragile techniques is easier than robust ones [29].

B. Robust steganography

In contrast of fragile techniques, bits manipulation of robust methods will not easily be removed from host file. Despite no method can guaranty that embedded data are not changeable but if amount of needed efforts for destructing information is considerable then it would be known as a robust method. To preserve security of embedded information its detection must be very hard [20].

Robust steganography is split into fingerprinting and watermarking. Application of fingerprinting is to put a mark on specific file for particular authorized customer. This specific mark can prove which customer has violated copyright law and is distributed particular copy of a file [31].

In opposite side of fingerprinting which aimed to protect the customers, watermarking preserves identity of file generator. Watermarking makes possible prosecution of people who hold illegal copies. Technically watermarking will be employed for mass production of CDs and DVDs, while fingerprinting is for especial copies.

When watermarking is not detectable and easily remove-able, it would be called imperceptible watermarking. However, there is visible version of watermarking to show particular patterns over an image. Example of this type of watermarking would be watermarks on some bank notes [3].

REFERENCES

[1] Mazdak Zamani, Azizah Bt Abdul Manaf, Rabiah Bt Ahmad, Farhang Jaryani, Saman Shojae Chaeikar and Hossein Rouhani Zeidanloo. "Genetic Audio Watermarking". International Conference on Recent Trends in Business Administration and Information Processing. March 26-27, 2010. Trivandrum, Kerala, India. Volume 70, Book: Information Processing and Management, Publisher: Springer Berlin Heidelberg, ISBN: 978-3-642-12213-2. 514-517.

[2] Saman Shojae Chaeikar, Azizah Bt Abdul Manaf and Mazdak Zamani. Comparative analysis between Master key and Interpretative Key Management (IKM) Framework to provide utilization guideline for researchers and developers. Cryptography and Security in Computing, ISBN: 978-953-51-0179-6. Publisher online InTech. 2012.

[3] Mazdak Zamani, Azizah Bt Abdul Manaf, Shahidan M. Abdullah, Saman Shojae Chaeikar. Mazdak Technique for PSNR Estimation in Audio Steganography. 2012 International Conference on Mechanical and Electrical Technology (ICMET 2012). July24-26, 2012, Kuala Lumpur, Malaysia.

[4] Mazdak Zamani, Hamed Taherdoost, Azizah A. Manaf, Rabiah B. Ahmad, and Akram M. Zeki. "An Artificial-Intelligence-Based Approach for Audio Steganography". MASAUM Journal of Open Probelms in Science and Engineering (MJOPSE). Volume: 1 Issue: 1 Month: October 2009. Pages 64-68.

[5] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, Farhang Jaryani, Hamed Taherdoost, Saman Shojae Chaeikar, and Hossein Rouhani Zeidanloo. "A Novel Approach for Genetic Audio Watermarking". Journal of Information Assurance and Security 5 (2010). ISSN: 1554-1010. Pages 102-111.

[6] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, Farhang Jaryani, Hamed Taherdoost, Saman Shojae Chaeikar, and Hossein Rouhani Zeidanloo. "Genetic Audio Steganography". International Journal on Recent Trends in Engineering & Technology [IJRTET], Volume 3, Issue 2, 2010. ISSN: 2158-5563. Pages 89-91.

[7] Shahidan M. Abdullah, Azizah A. Manaf, and Mazdak Zamani. Recursive Reversible Image Watermarking Using Enhancement of Difference Expansion Techniques. Journal of Information Security Research. Volume 1 Number 2. June 2010. Pages 64-70.

[8] Mazdak Zamani, Azizah Abdul Manaf, Hossein Rouhani Zeidanloo, and Saman Shojae Chaeikar. "Genetic Subtitution-Based Audio Steganography for High-Capacity Applications". International Journal for Internet Technology and Secured Transactions (IJITST). Pages 97-110. ISSN 1748-569X (Print), ISSN: 1748 - 5703 (Online). Volume 3 Issue 1, April 2011. Inderscience Publishers, Geneva, Switzerland.

[9] Akram M. Zeki, Azizah A. Manaf, Adamu A. Ibrahim and Mazdak Zamani. A Robust Watermark Embedding in Smooth Areas. Research Journal of Information Technology. Year: 2011. Volume: 3. Issue: 2. Page No.: 123-131.

[10] Mazdak Zamani, Azizah Bt Abdul Manaf, Shahidan M. Abdullah. An Overview on Audio Steganography Techniques. International Journal of Digital Content Technology and its Applications (JDCTA). Published by AICIT (Advanced Institute of Convergence Information Technology). 2012.

[11] Maziar Janbeglou, Mazdak Zamani, Suhaimi Ibrahim. Improving the Security of Protected Wireless Internet Access from Insider Attacks. Advances in information Sciences and Service Sciences (AISS). Volume4, Number12, July 2012.

[12] Mojtaba Alizadeh, Mazdak Zamani, Ali Rafiei Shahemabadi, Jafar Shayan, Ahmad Azarnik. A Survey on Attacks in RFID Networks. Open International Journal of Informatics (OIJI). Vol 1 (2012).

[13] Mazdak Zamani, Azizah Abdul Manaf, and Rabiah Ahmad. "Knots of Substitution Techniques of Audio Steganography". The 2009 International Conference on Telecom Technology and Applications. Pages 415-419. June 6-8, 2009. Manila, Philippines.

[14] Mazdak Zamani, Azizah Abdul Manaf, and Rabiah Ahmad. "Current Problems of Substitution Technique of Audio Steganography". The 2009 International Conference on Artificial Intelligence and Pattern Recognition. ISBN: 978-1-60651-007-0. Pages 154-160. 13-16 July 2009. Orlando, Florida, USA.

[15] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, Akram Zeki, and Shahidan Abdullah. "Genetic Algorithm as an Approach to Resolve the Problems of Substitution Techniques of Audio Steganography". The 2009 International Conference on Genetic and Evolutionary Methods. 13-16 July 2009.. Pages 170-175. Las Vegas, Nevada, USA.

[16] Mazdak Zamani and Azizah Abdul Manaf. "Azizah's Formula to Measure the Efficiency of Steganography Techniques". 2nd International Conference on Information and Multimedia Technology (ICIMT 2010). December 28-30, 2010. Hong Kong, China.

[17] Mazdak Zamani and Azizah Abdul Manaf. "Mazdak's Method to Estimate the PSNR of Audio Steganography Techniques". International Conference on Computer and Computational Intelligence (ICCCI 2010). December 25-26, 2010. Nanning, China.

[18] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, Akram Zeki, and Shahidan Abdullah. "A Genetic-Algorithm-Based Approach for Audio Steganography". International Conference on Communities and Communications. World Academy of Science, Engineering and Technology 54 2009. ISSN: 2070-3740. Pages: 359-363. 24-26 June 2009. Paris, France.

[19] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, and Akram Zeki. "An Approach to Improve the Robustness of Substitution Techniques of Audio Steganography". 2nd IEEE International

Conference on Computer Science and Information Technology 2009. Volume 2. Pages 5-9. 8 - 11 August 2009. Beijing, China.

[20] Mazdak Zamani, Hamed Taherdoost, Azizah Abdul Manaf, Rabiah Ahmad, and Akram Zeki. "Robust Audio Steganography via Genetic Algorithm". Third International Conference on Information & Communication Technologies ICICT2009. ISBN: 9781424446087. Pages 149 - 153. 15-16 August 2009. Karachi, Pakistan.

[21] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, Akram Zeki, and Pritheega Magalingam. "A Novel Approach for Audio Watermarking". Fifth International Conference on Information Assurance and Security. ISBN: 978-0-7695-3744-3. Pages 83-86. 18-20, August 2009. Xi'an, China.

[22] Mazdak Zamani, Azizah Abdul Manaf, Rabiah Ahmad, Farhang Jaryani, Hamed Taherdoost, and Akram Zeki. "A Secure Audio Steganography Approach". The 4th International Conference for Internet Technology and Secured Transactions. ISBN: 978-0-9564263-1-4. Pages 501-506. 9-12 November 2009. London, UK.

[23] Saman Shojae Chaeikar, Shukor Abd Razak, Shohreh Honarbakhsh, Hossein Rouhani Zeidanloo, Mazdak Zamani and Farhang Jaryani. Interpretative Key Management (IKM), A Novel Framework. 2010 International Conference on Computer Research and Development (ICCRD 2010). May 7 - 9, 2010. Kuala Lumpur, Malaysia.

[24] Akram M. Zeki, Azizah A. Manaf, and Mazdak Zamani. Bit-Plane Model: Theory and Implementation. Engineering Conference 2010 (EnCon2010). 14-16 April 2010. Kuching, Sarawak, Malaysia.

[25] Nuha Omran Abokhdair, Azizah Bt Abdul Manaf, Mazdak Zamani. Integration of Chaotic Map and Confusion Technique for Color Medical Image Encryption. 6th International Conference on Digital Content, Multimedia Technology and its Applications (IDC2010). 16-18 August 2010. Seoul, Korea.

[26] Shahidan M. Abdullah, Azizah A. Manaf, and Mazdak Zamani. Capacity and Quality Improvement in Reversible Image Watermarking Approach. 6th International Conference on Networked Computing and Advanced Information Management. 16-18 August 2010. Seoul, Korea.

[27] Mazdak Zamani, Azizah Abdul Manaf, and Rabiah Ahmad. "Knots of Substitution Techniques of Audio Steganography". 2009 International Conference on Computer Engineering and Applications. IPCSIT vol.2 (2011) IACSIT Press, Singapore. Pp 370-374.

[28] Mazdak Zamani, Azizah Bt Abdul Manaf, Shahidan M. Abdullah, Saman Shojae Chaeikar. Correlation between PSNR and Bit per Sample Rate in Audio Steganography. 11th International Conference on Signal Processing (SIP '12). Pages 163-168. Saint Malo & Mont Saint-Michel. France, April 2-4, 2012.

[29] Mazdak Zamani, Azizah Bt Abdul Manaf, Shahidan M. Abdullah. Correlation between PSNR and Size Ratio in Audio Steganography. 11th International Conference on Telecommunications and Informatics (TELE-INFO '12). Pages 82-87. Saint Malo & Mont Saint-Michel. France, April 2-4, 2012.

[30] Mazdak Zamani, Azizah Bt Abdul Manaf, Shahidan M. Abdullah. Efficient Embedding for Audio Steganography. 2nd International Conference on Environment, Economics, Energy, Devices, Systems, Communications, Computers, Mathematics (EDSCM '13). Pages 195-199. Saint Malo & Mont Saint-Michel. France, April 2-4, 2012.

[31] Mazdak Zamani, Azizah Abdul Manaf, and Rusni Daruis. "Azizah Technique for Efficiency Measurement in Steganography". 8th IDCTA: 2012 International Conference on Digital Content, Multimedia Technology and its Applications. June 26 - 28, 2012. Jeju Island, Korea.

[32] Mojtaba Ali Zadeh, Mazleena Salleh, Mazdak Zamani, Jafar Shayan, Sasan Karamizadeh. "Security and Performance Evaluation of Lightweight Cryptographic Algorithms in RFID". 16th WSEAS International Conference on Communications (part of the 16th CSCC / CSCC 2012). Kos Island, Greece. July 14-17, 2012.

[33] Mojtaba Ali Zadeh, Mazdak Zamani, Jafar Shayan, Touraj Khodadadi. "Code Analysis of Lightweight Encryption Algorithms Using in RFID Systems to Improve Cipher Performance." The 2012 IEEE Conference on Open Systems. Kuala Lumpur, Malaysia. 21st – 24th October 2012.

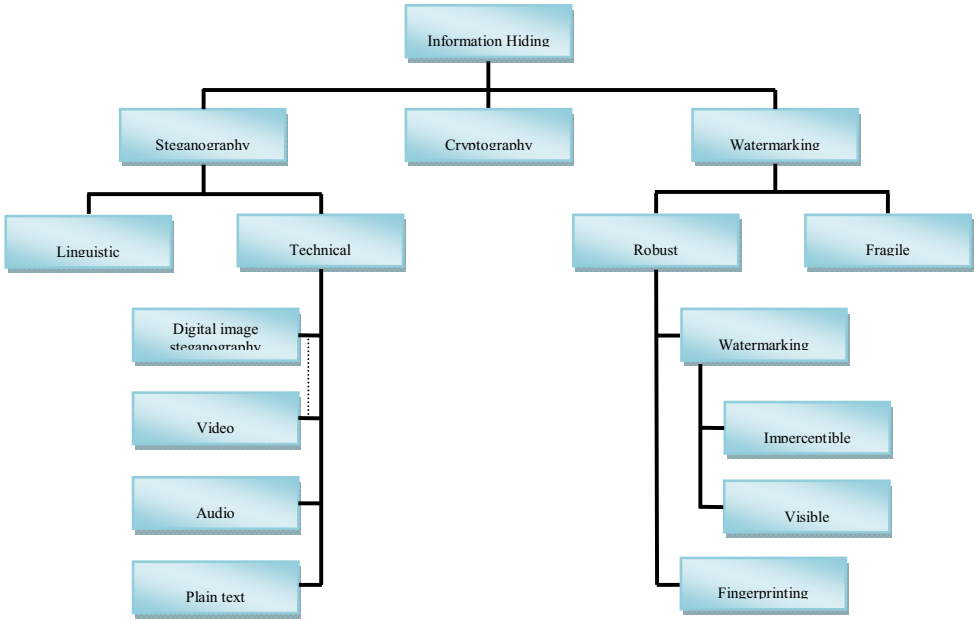

Figure 2. Classification of data hiding techniques