ASTON UNIVERSITY

WATERMARKING BIOMEDICAL TIME SERIES DATA

BASAVA RAJESWARI MATAM

A thesis submitted for the degree of Doctor Of Philosophy, 2009

Thesis Summary

This thesis addresses the problem of information hiding in low dimensional digital data focussing on issues of privacy and security in Electronic Patient Health Records (EPHRs). The thesis proposes a new security protocol based on data hiding techniques for EPHRs. This thesis contends that embedding of sensitive patient information (such as personal textual information exemplified by name or other personal identifiers such as genetic information) inside the EPHR is the most appropriate solution currently available to resolve the issues of security in EPHRs. Watermarking techniques are applied to one-dimensional time series data such as the electroencephalogram (EEG) to show that they add a level of confidence (in terms of privacy and security) in an individual's diverse bio-profile (the digital fingerprint of an individual's medical history), ensure belief that the data being analysed does indeed belong to the correct person, and also that it is not being accessed by unauthorised personnel.

Embedding information inside single channel biomedical time series data is more difficult than the standard application for images due to the reduced redundancy. A data hiding approach which has an in built capability to protect against illegal data snooping is developed. The capability of this secure method is enhanced by embedding not just a single message but multiple messages into an example one-dimensional EEG signal. Embedding multiple messages of similar characteristics, for example identities of clinicians accessing the medical record helps in creating a log of access while embedding multiple messages of dissimilar characteristics (such as sensitive patient information requiring a high level of security, clinician's notes requiring a moderate level of security and tamper detection and authentication code requiring a low level of security) into an EPHR enhances confidence in the use of the EPHR.

The novel method of embedding multiple messages of both similar and dissimilar characteristics into a single channel EEG demonstrated in this thesis shows how this embedding of data boosts the implementation and use of the EPHR securely.

Keywords: Biomedical Time Series, Electronic Patient Health Records, Patient Privacy, Security, Watermarking

Contents

List of Figures

List of Tables

Notation

Bold face capital letters ($\mathbf{S}$) denote matrices. Bold face small letters ($\mathbf{s}$) represent vectors. Matrices/vectors with a check ($\check{\mathbf{S}}/\check{\mathbf{s}}$) represent the unknown processes. Matrices/vectors with a Bar ($\bar{\mathbf{S}}/\bar{\mathbf{s}}$) represent the perturbed version of the corresponding matrix/vector. Matrices in boldface ($\mathbf{S}$) with no accents represent the visible processes. Matrices/vectors with a tilde ($\tilde{\mathbf{S}}/\tilde{\mathbf{s}}$) represent data with embedded information (watermarked). And finally matrices/vectors with a hat ($\hat{\mathbf{S}}/\hat{\mathbf{s}}$) represent watermarked data after an attack.

$\mathbf{B}_{ij}$	Index denoting the structure order of sample $\mathbf{s}(i)$ compared with sample $\mathbf{s}(j)$ of a time series
N_{ts}	Length of a time series
$\mathbf{SO}$	Structure order matrix of a time series
$v(i)$	sample whose co-ordinates are the rank order values of two time series $\mathbf{s}_a$ and $\mathbf{s}_b$, $(\mathbf{r}_a(i), \mathbf{r}_b(i))$
$\mathbf{r}_{spec}$	Rank order of the spectrum of time series $\mathbf{s}$
N_{clust}	Number of clusters of estimated sources
$corr$	Correlation between two time series
$\mathbf{w}$	Independent component, row of $\mathbf{W}$
ζ	Threshold for η. $\eta > \zeta$, cover work is destroyed
$\mathcal{H}$	Output of decision, $\mathbf{m}$ can/cannot be recovered from $\hat{\mathbf{c}}$
ξ	Perturbation applied to the set of observations
$\mathbf{Pr}$	Probability distribution
$\mathbb{R}$	Real valued numbers
$\bar{\mathbf{X}}$	Perturbed observation matrix
$\bar{\mathbf{S}}$	Sources estimated from $\bar{\mathbf{X}}$ as input
$\bar{\mathbf{W}}$	Estimated separating matrix from $\bar{\mathbf{X}}$
$\bar{\mathbf{A}}$	Estimated mixing matrix from $\bar{\mathbf{X}}$
E	Perturbation matrix
φ	A small perturbation of $\check{\mathbf{A}}$

$\mathbf{B}$	Bandwidth of input signal
$\mathbf{c}$	Original unwatermarked document also called cover
$\tilde{\mathbf{c}}$	Watermarked cover
$\hat{\mathbf{c}}$	Watermarked cover distorted due to signal processing attack
$\mathbf{m}$	Message to be embedded, watermark
WM	Binary representation of $\mathbf{m}$
$dist$	$\sum[WM \oplus \widehat{WM}]$
D_{Emb}	$\mathbf{c}$ - $\tilde{\mathbf{c}}$
η	Attack (signal processing/malicious attack on $\tilde{\mathbf{c}}$
$\mathbf{K}$	Vector of possible samples of $\mathbf{c}$ that can be watermarked.
$\mathbf{k}$	key (random selection of samples to be watermarked), $\mathbf{k} \in \mathbf{K}$
$\hat{\mathbf{m}}$	Estimate of embedded message retrieved at decoder
$Nsamp$	Length of $\mathbf{c}$
p_i	Probability distribution of $\tilde{\mathbf{c}}$
q_i	Probability distribution of $\mathbf{c}$
$\mathcal{R}$	Rate of information (Length of WM)
y	Output of Decoder
$\mathcal{F}$	Watermark embedding technique
ϑ	Acceptable level of $\mathcal{D}_{Emb}$
$\mathcal{KL}$	Kullback Leibler divergence
ρ	Error threshold below which a positive identification of the existence of a watermark is considered
$\mathcal{T}$	Transform applied to $\mathbf{c}$
N_{nov}	Length of non-overlapping segments of $\mathbf{c}$
N_{ov}	Length of overlapping segments of $\mathbf{c}$
F	Frequency spectrum of $\mathbf{c}$
$\mathbb{E}$	Expectation
μ	Mean of a signal
s	Scale factor of wavelet
τ	Translation parameter of wavelet
$diff$	Vector of difference between consecutive sample of C which have been sorted
$\mathbf{o}$	Dither signal used in DM_QIM

ψ	Mother wavelet, Haar
C_a	Approximate co-efficients of wavelet decomposition of $\mathbf{c}$
C_b	Detail co-efficients of wavelet decomposition of $\mathbf{c}$
$\mathbf{Y}$	Eigen vectors obtained from principal component analysis of $\mathbf{c}$
$\mathbf{X}$	Input matrix to PCA and ICA constructed from 1D $\mathbf{c}$
$\tilde{\mathbf{X}}$	Watermarked $\mathbf{X}$
$\acute{C}$	Principal components decomposition of $\hat{\mathbf{X}}$
$\check{\mathbf{S}}$	Matrix of unknown statistically indpendent sources that are linearly combined to obtain the observation vectors
$\mathbf{S}$	Matrix of estimated sources from $\mathbf{X}$ using the ICA
$\tilde{\mathbf{S}}$	Matrix of sources containing both watermarked and non-watermarked sources
l	Number of estimated independent sources
p	Number of observation vectors
$\check{\mathbf{A}}$	Unknown mixing matrix
$\mathbf{d}$	Delay vector of observation vector
EmbWin	Delay embedding window size
f_s	Sampling frequency of signal
f	frequency of the slowest moving component of a mixed signal
$\mathbf{A}$	Mixing matrix estimated by the ICA
$\mathbf{W}$	Separating matrix estimated by the ICA
$\mathbf{x}$	Observation vector
$\mathbf{s}$	An estimated source
$\mathbf{s}_{wm}$	Source to be watermarked
δ	Quantisation index used to modulate samples of C with the watermark
$\mathbf{N}_c$	Length of C
$\mathbf{N}_{wm}$	Length of the watermark
N	Order of the LPC filter
A	Linear prediction polynomial of order N
e_p	Error between p^{th} sample of original and reconstructed signal using LPC
S	Size of a bin in the histogram of the samples of C
$\mathbf{B}$	Total number of bins
N_b	Number of samples in each bin S
Δ	Vector lattice grid on which the samples of C lie
d	Step size of one of the quantisation levels of Δ

S_{γ}	Size of bin in the histogram of the samples of $diff$
$\mathsf{N}_{b_{\gamma}}$	Number of samples of $diff$ in each bin of size S_{γ}
ε_c	Distortion to $\tilde{\mathbf{c}}$ due to η
ε_C	Distortion to $\tilde{C}$ due to distortion of $\tilde{\mathbf{c}}$
$\check{\delta}$	Different δ values assumed to obtain an estimate of the true δ and $\mathbf{k}$
$\mathcal{N}(0,1)$	Normal distribution of zero mean and unit variance
ε	Estimate of ε used to find the probable value of δ and $\mathbf{k}$
$\mathbf{m}_{sim_i}$	Similar characteristics messages
υ	Value of $\check{\delta}$ after which the length of the estimate of $\mathbf{k}$ by the attacker remains constant
d	Distance between two time series
$\mathbf{r}$	Rank order vector of a time series

Acknowledgements

I thank Professor David Lowe for giving me the opportunity to work with him. His guidance and tireless effort to teach me is the reason I have been able to complete this work. All the work I have been able to produce in the last three years is largely due to him. His encouragement and confidence in my capability gave me the strength to present and discuss my work with my peers.

I would also like to thank the entire staff of NCRG for the help I have received at various times in the years I have spent here. A special thanks to Vicky Bond who was more a friend than a member of staff. A big thankyou to Kanchan for being a friend and filling up the coffee machine with more chocolate.

I take this opportunity to thank some of the important people who supported me the last three years Celestin Pham, Pierre Beguire, Jack Raymond, Michel Randrianandrasana, Erik Casagrande, Steve d'Aguiar, Mingmanas Sivaraksa, Thomas Bermudez, Diar Nasiev, Anthony Jones, Michael Vrettas, Alex Boukouvalas, Steve Dalton, Remi Barillec, Harry Goldingay, Yuan Shen, Roberto Castro Alamino, Andras Joo, Hannah and James. Many thanks to Dharmesh Maniyar, Boremi Toch, Alex Brulo and Bishnu Lamichhane.

Lastly I thank my family who are the most important people in my life. Everything I have achieved and everything I am is due to the unconditional love and support of my mother, Sunila Rajan and the rest of my family Lakshmi Devi, Kumar, Rajyalakshmi, Bindu Madhavi, Kavitha and Shiva Prasad. I would also like to thank my two closest friends Satyabhama and Nayana for just being there when I needed a friend.

1 INTRODUCTION

CONTENTS

This thesis addresses the issue of a technical solution towards the problem of patient privacy in the use of computerised biomedical records. The approach adopted in this thesis is based on watermarking. As defined in [47] watermarking extends the information of the cover (data used to embed information). Bender [12] states that 'data hiding, embedding information into digital media for the purpose of identification, annotation and copyright is a form of steganography (hiding information in plain sight)'. Cox et al [24] state that 'though the applications and requirements for steganography and watermarking may be different, the actual techniques used for watermarking and steganography may be very similar, or in some cases identical'. Based on these definitions the work presented in this thesis is termed as watermarking (data embedding) though the line between steganography (data hiding) and watermarking is thin and overlap exists. Current security protocols for EPHRs are designed based on Information Communication Technologies (ICT) such as smart cards and logical deletion (the practice of marking data as being no longer applicable [90]). In role based access systems wherein only parts of a health record are available to the different personnel of a hospital, the cryptographic solution to hiding patient-sensitive data is inherently insecure. Since the personal information has to be decrypted at some point to identify that the non-personal biomedical data is indeed related to a particular patient, accidental leakage of data is possible.

The solution provided in this thesis is based on data embedding methods which are complementary to cryptography. It will be shown how the novel data embedding approach implemented in this thesis for EPHRs has the capability to provide a viable solution to a host of problems impeding the implementation of EPHRs. Patient privacy is secured by means of hiding the personal details of the patient in an EPHR. This is a steganographic watermark. Cox et al [24] define a steganographic watemark as embedded information that is related to the content but is hidden. The sensitive personal information is the secret message. It relates a patient with a particular EPHR but it is intended for only authorised health providers (clinicians in direct contact with the patient). Its existence is concealed from the other health personnel such as those involved with billing and maintainance whose work could be conducted by using an identifier. Additional information pertaining to information about the EPHR (for example the hospital/clinician who conducted the tests or, details of how the tests were conducted which could provide additional information to the clinician for diagnostic purposes) will be embedded into the EPHR using watermarking techniques. This information could be used to tag the EPHR such that researchers and adminstrative personnel can locate the EPHR with the help of this tag if required. This information is defined as a non-steganographic watermark or simply watermark in [24]. In this thesis though some of the embedded information belongs to the class of steganographic watermarks based on its use, for simplicity all the embedded data will be referred to as watermarks and the data hiding technique for both steganographic and non-steganographic watermarks will remain the same. The watermarking method will be tested and validated as a possible solution to some of the drawbacks of the standard ICT based security methods for EPHRs.

In this chapter the EPHR, its advantages and the issues impeding its rollout are discussed in section I. Watermarking methods are introduced in section II. And lastly a summary of the benefits that data embedding technologies could bring to EPHRs and help to resolve the issues of patient privacy in the use of EPHRs is presented.

1.1 Electronic Patient Health Record (EPHR)

Computerised medical records and the advances in telecommunication infrastructure enable remote health care services, where treatment without the physical presence of a clinician is possible. EPHRs, capable of providing personalised health care, are being investigated and are in varying stages of implementation. Detailed descriptions of EPHR implementations can be found in [32, 85, 4, 22, 38, 17, 69]. The advantages of an EPHR include remote health care, faster access to the health record resulting in quicker diagnosis and efficient treatment in case of an emergency, better prescription services, and prevention of duplicating medical tests. Researchers believe EPHRs can provide the infrastructure to permit individualised treatment to every patient. This is possible by applying predictive and decision support models to a Bioprofile (an electronic record of an individual's health progression from birth to death possibly including his/her genetic profile along with a personal identifier containing the name, sex, age, and address of the individual, which relates the bioprofile to its owner), and studying the results of automated machine learning to provide prognosis/diagnosis.

The above advantages can be fully gained only when all EPHRs/bioprofiles are centrally accessible by clinicians and researchers for the purpose of diagnosis and study. However a central database with every individual's health record containing sensitive and private information is highly vulnerable to abuse, increasing the risk of loss of privacy and distress to the patient. This is evident from the reports on data loss from hospitals [11]. These reports demonstrate the potential risk of psychological and material damage to the patient if personal details are easily accessible to anybody who gains access to the eHealth network. A major bottleneck in implementing the EPHR fully comes from partially valid patient fears on questions such as: 'Could fund-raisers get details of individuals suffering from a particular disease to approach for donations? Could employers obtain private health records to reduce their workers compensation costs, or identify employees who may be costly in the future? Could computer hackers release the information onto the Internet for everyone to see? How will patients be able to control access to, or find out who has viewed their medical records?'

Secure and error-free storage and transmission of the EPHR/bioprofile has therefore become vital to gain public confidence and acceptance of the eHealth system [52]. There are two other major issues impeding the roll out of the EPHRs. One, there is no interoperability between different hospital networks and, two, no uniform standards for storing different medical data exist.

The health care system depicted in Figure 1.1 is an example computerised healthcare system. The different health care units are linked together by telecommunication networks. As can be seen from figure 1.1, patients registered in a local hospital but requiring treatment in a large hospital need not repeat medical tests conducted in the local hospital. Since all the healthcare units are connected to each other and the EPHRs stored in a central database, records from any healthcare unit can be accessed by any other healthcare unit in real time. The advantages of such health care units are significant so, in this thesis a design of an eHealth system that can benefit from the advantages and simultaneously curtail the risks involved is presented. The security protocols implemented in most clinical standards approved by governments of various countries include [30, 49]:

- Logging in to a health centre database with a smart card and Personal Identification Number (PIN),
- Role based access (access to parts of the medical document depending on the role of the health service professional),
- The right of patients to determine what medical reports and data are saved in the medical document, and who is allowed to access the medical document,

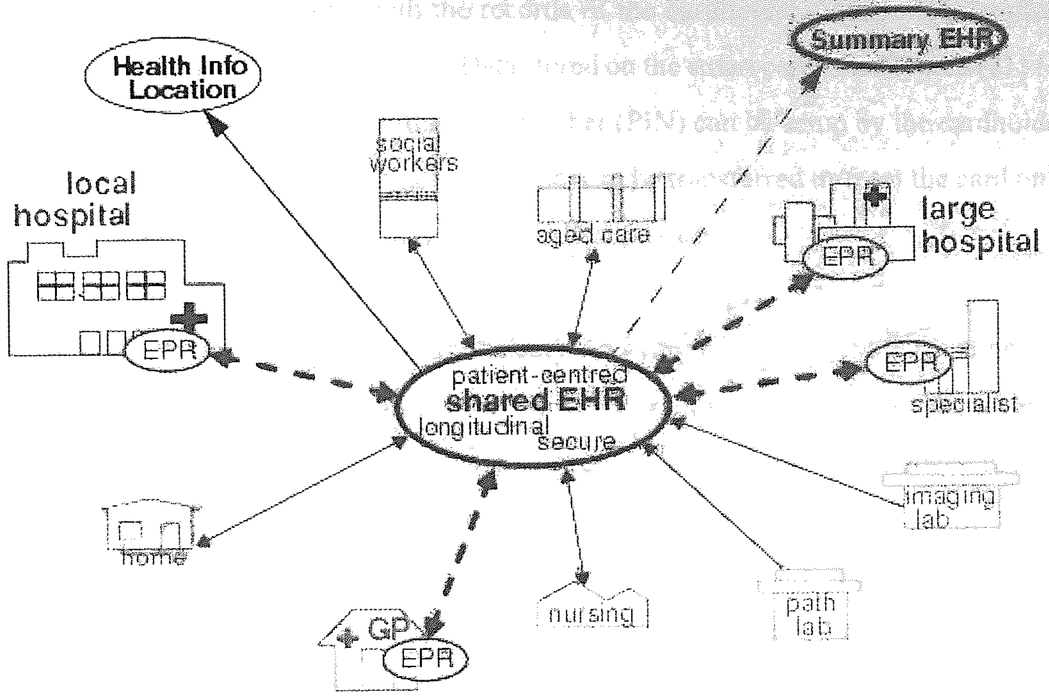


Figure 1.1: The design of an EPHR system from openEHR termed 'community shared-care context'. This picture depicts a fully connected healthcare services system on a probable regional level.

- Encryption of the personal details by standard crytography such as RSA.

Some examples of eHealth systems and eHealth standards focussing on the security measures implemented in them are presented in the next section. The drawbacks and limitations of the above security measures and the need for a method based on data hiding techniques to counter these issues is also discussed.

1.1.1 Examples of EPHR implementations

Though EPHRs are being tested and implemented in different countries as already mentioned, some examples of extensive deployment of EPHRs include the models in Taiwan and Estonia.

Taiwan:

The Taiwanese smart card-based health IC card [22] is designed to be a mobile data carrier held by the patient. Its personal information section carries the card number and date of issuance in addition to the cardholder's name, gender, date of birth, ID number and photo. Its health insurance related information section further registers major diseases, the number of visits and admissions to medical institutions, the last menstruation period and pregnancy examinations, along with the records of the cardholder's insurance premium and accumulated medical expenditures. Data stored on the smart card is encrypted [21] for security purposes. A personal identification number (PIN) can be setup by the cardholder to protect the information on the smart card. Data can be transferred to/from the card only after a strict authorisation and mutual authentication process.

Estonia:

The National Identity Card which is compulsory for every Estonian has the usual person identifying features - name, picture, date of birth and personal code [49, 86]. The security of the information on the card is based on the personal identification code enabled in each card and a certificate in the ID-card which enables digital signing.

These example eHealth systems rely on ICT based solutions for security. The health IC card contains a complete description of an individual's health which is personal and sensitive. The security measures based on ICT have largely been tested for the banking industry. Though these measures provide secure transactions between a single user and a database, they have limitations when used in a multiuser federated database environment such as an eHealth network. The drawbacks of ICT based security protocols for eHealth systems is discussed in section 1.1.3

1.1.2 Examples of EPHR standards

Different standards are being developed for computerising medical records. Some of the standards currently under use and development are DICOM, Health Level 7 (HL7) Clinical Development Architecture (CDA)[HL7 CDA Release 2.0 2005], CEN EN 13606 EHRcom [CEN prEN 13606-1 2004], and openEHR , and have been described in [30]. Health networks of different countries and different hospitals in a country are designed using different standards. The disadvantage of multiple standards is that there is little or no interoperability. Also, some of the standards being proprietary impedes small hospitals in using them, making interoperable national health networks difficult.

Two of the standards for EPHRs being used extensively are openEHR and the protocols designed by 'Integrating the Health Enterprise (IHE), a joint initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information' [43]. The security measures implemented in these two standards are discussed below with their disadvantages.

The security policy of the openEHR standard [73] is as shown in figure 1.2.

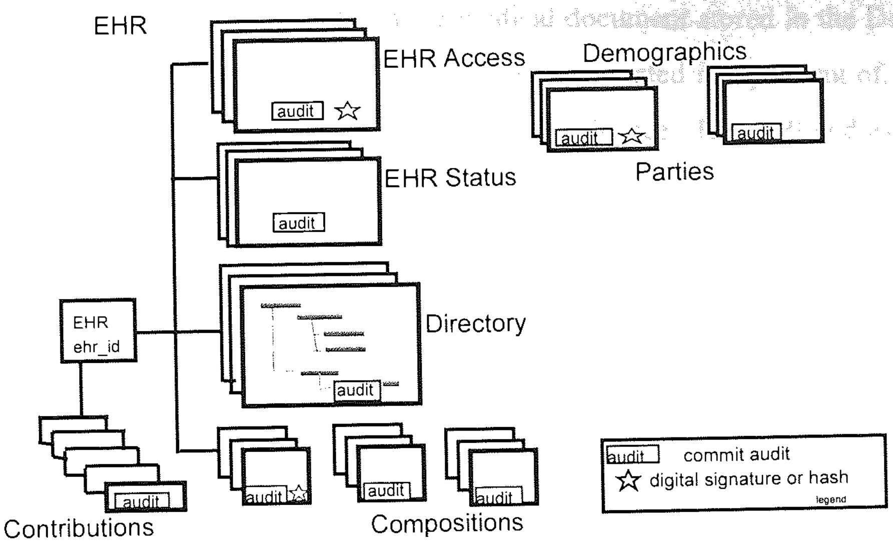


Figure 1.2: Security features of the openEHR depicting the separation of the personal and medical information in an EPHR. It also shows the various frames in a probable EPHR.

The general features of the security policy of the openEHR mainly include indelibility, audit trailing and anonymity. Since the information in a health record cannot be deleted, access to parts of the health record by the different health personnel is achieved by marking the data in such a way as to make it appear deleted. Any access of a health record by a health care professional is trailed with user identity, time-stamp, reason, optionally digital signature and relevant version information. Another security feature implemented in the openEHR is the separation of the personal and medical information in an EPHR as depicted in figure 1.2. The eHealth system is configured such that theft of the EPHR does not provide any clue to the identity of the patient. A cross-reference database protected by means of encryption or other security mechanisms is used to relate the EPHR to a demographic file.

Versioning in the openEHR is its most basic and important security related feature for data integrity. All logical changes and deletions as well as additions are physically implemented as new Versions rather than changes to existing information items. The openEHR also states that there exists a possibility to digitally sign each Version. The security mechanisms are left unspecified to be decided by the third party vendor who implements the network.

The Cross-Enterprise Document Sharing framework (XDS) designed by the IHE [43] to provide remote access to clinical documents is shown in figure 1.3. The hospital, denoted as the Document Source, generates a unique Electronic Patient Index (EPA-I) for each patient's medical document and the medical document stored in the Document Repository. A Master Patient Index (MPI) number is generated independent of, and unknown to the Document Source by the Patient Identity Source. The MPI and associated EPA-I are stored in the Document Registry. The Document Consumer (hospital or clinician) requiring access to a patient's EPHR queries the Document Registry with knowledge of the EPA-I to obtain the MPI. The MPI is released after verification of the source requiring access and the Document Consumer with knowledge of the MPI can access and view the medical record.

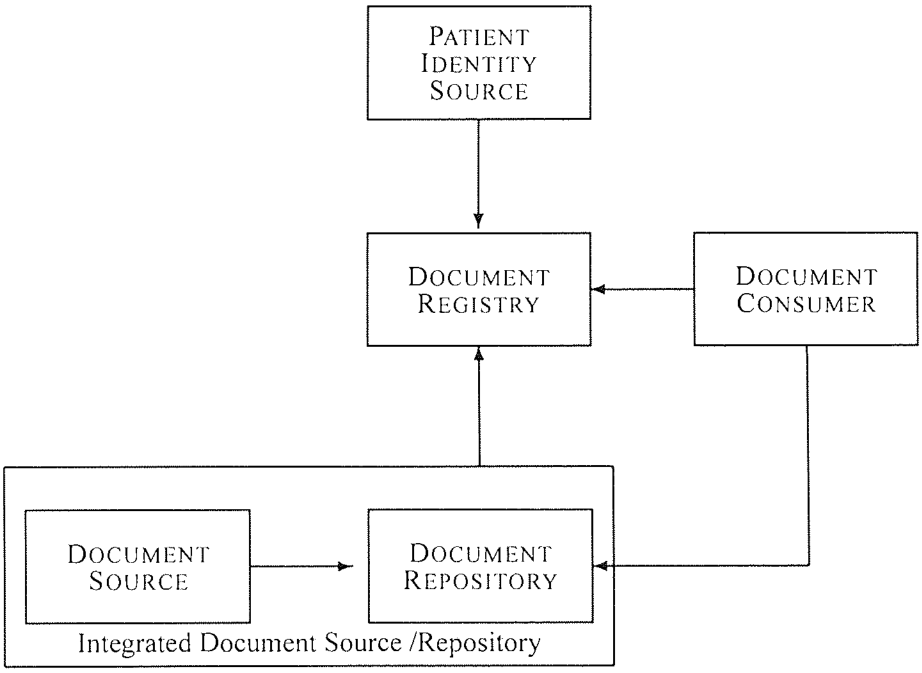


Figure 1.3: An eHealth system framework designed by the IHE, 'Cross enterprise document sharing diagram'. Figure shows the centally accessible EPHR database contained in document repository where each EPHR is identified by a ten digit index. The ten digit index is generated independent of the hospital (document source) producing the EPHR by the patient identity source. The document registry provides the link between the EPHR stored and the identification number needed to access the EPHR.

The security policy of the openEHR standard is suitable to prevent unauthorised access of the EPHR but it does not provide a mechanism to prevent accidental leakage of information from the EPHR by authorised personnel. Secondly security protocols in the openEHR based EPHR standard being implemented by the third party vendors creates a bottleneck to fully connect health networks designed by different vendors. Similarly the XDS framework based eHealth networks are protected from unauthorised access but not against loss of data by authorised personnel. The examples of computerised healthcare systems of Taiwan and Estonia denote a small selection of countries using smart card based technologies for their respective EPHR systems. Similarly the list of the standards for EPHRs is representative only. The use of smart cards for health and EPHRs are advantageous in the sense that they provide strict authorisation procedures but the risk of misuse also exist. One example is that of clinicians logging in, and leaving it open for other clinicians to access records in busy A&E departments [1]. The argument in favour of sharing smart cards is that 'precious time is saved and patients can be treated more efficiently, and the systems are placed in rooms with limited access'. This method is insecure in that health personnel such as adminstrators, maintainance and not necessarily clinicians with access to the rooms could also access the health records.

1.1.3 Security loopholes in the EPHR standards

Security mechanisms utilising smart card technology as defined in the security policy of standards such as openEHR have drawbacks when used in large-scale interconnected networks with multiple users of the networks having the same access privileges to the same set of data. It is also visible to a malicious attacker who can easily detach it and attach new data to the medical record. This form of attack is difficult to notice and can lead to complications for the patient.

If logging in each time is not mandatory, or smart cards are shared by clinicians in a real time environment such as an Emergency Department of a hospital, patient identity and security of the sensitive medical data can be heavily compromised. An intruder gaining illegal access to patients' records can neither be prevented nor traced. Secondly the patient index or identifier is a set of characters and numbers. An erroneous entry of this identifier cannot be identified at the receiving hospital. Lastly the identifier is a header attached to the medical document. The link between medical data and patient details can occasionally get mangled by protocol converters [54].

Complete anonymisation or deletion of the details such as name, address from medical files, but inclusion of information such as age, gender, ethnic origin, demographics relevant for the study being conducted is the norm in research [3]. The probability of relating an individual to a medical record or a particular group of individuals to a medical condition is significant even in such 'de-identified' databases as shown by [72]. Also this method of anonymisation is not suitable for medical records used for diagnostic purposes in a health centre.

Hence security measures being left open to be implemented by third party vendors as in the openEHR or any medical data standard may not be adequate. Security measures built on security protocols mentioned above may thus result in deadlocks. In brief, techniques relying on smart cards and identifiers for privacy and security in EPHRs fail in that they cannot ensure 'the received medical data is both legally and medically appropriate'.

Therefore, an alternative perspective is needed which can address some of the legal and ethical issues, but not suffer the weaknesses of current security implementations of the EPHR.

1.2 Watermarking

In this section, an introduction to watermarking, its applications and the advantages data hiding methods could provide to secure and protect the privacy of the patient in an EPHR is presented. A demonstration of how watermarking principles could help in preventing accidental leakage of patient sensitive information is given. It will also be shown how the hidden information could help in linking a medical record to a particular patient and provide a log of access (identify personnel who have accessed the EPHR) of the EPHR.

The need for secure communications and the tools used, cryptography (data encryption) and steganography (data hiding) can be traced back to a period between 600BC and 400BC [23]. Cryptography renders messages unintelligible to unauthorised persons who intercept them while steganography conceals the message itself from unauthorised persons [78]. From [24] and the references within, paper watermarks were first created around 1282. The purpose of the early watermarks is not known but by the eighteenth century the use of watermarks as trademarks become common.

The use of watermarks as hidden messages about the content in which they are embedded is the most general and popular definition of watermarks [24]. This property of watermarking (concealing information which needs to be secured in other data) can provide both anonymity to the embedded information and authentication of the cover work (data used to hide the information) [24].

One of the earliest examples of watermarking is a message hidden as the first letter of each chapter of a book, Hypnerotomachia Poliphili, published in 1499 [24]. The process of modifying cover data representing images, audio or video to contain hidden information is called watermarking [48]. Various watermarking or data hiding techniques for text, video, audio, image and 3D signals have been designed, attacked and countered [26]. With the use of appropriate embedding/detecting techniques the hidden information should be recoverable even if the host signal is compressed, edited or converted from digital to analog format and back. The embedded data can be used in applications requiring authentication and tamper detection of the cover data at the receiver. From these observations the need for watermarking techniques becomes apparent with growing concerns of digital piracy and authentication.

1.2.1 Data embedding techniques -applications in an eHealth scenario

Following is a brief list of some of the applications of digital data embedding methods [24, 78] as applied to multimedia data. A modification and adapatation of these methods to the biomedical domain is given below.

- Passive and active copyright protection: Digital watermarking provides a means to identify the owner or distributor of digital data by embedding important control, descriptive or reference information in a given work. This application is mainly used to prevent the unauthorised copying of digital data via the internet.

  The capability of the watermark to withstand changes as the cover is modified gives a measure of its strength to survive further processing of the watermarked cover. A watermark is said to be robust if it is distorted only when the cover work is destroyed to a large extent. A fragile watermark is one which distorts when the watermarked cover undergoes any slight changes while a semi-fragile watermark can tolerate distortions to the cover due to unintentional attacks/signal processing of the watermarked cover. Semi-fragile watermarks can also be used to detect non-malacious/malicious tampering. Watermarks used for copyright protection need to be robust against any combination of signal processing attacks.

  In an EPHR where patient identification details should not be available to unauthorised accessors of the EPHR, the personal details of the patient take the preference of a semi-fragile steganograhic watermark. The personal details are a means of identifying the rightful owner of an EPHR. Hence they need to be robust against a variety of unintentional signal processing transformations of the cover.

- Broadcast monitoring: A computer monitors broadcasts and compares the received signals with a database of known works. Passive monitoring systems try to directly recognise the content being broadcast while active monitoring systems rely on associated information (watermarks) that is broadcast along with the content.

  EPHRs contain sensitive information which when broadcast (or accessed indiscriminately) have the potential to cause distress to the patient. A means of regulating the access of the EPHR and logging the identities of the accessors is important. The application of broadcast monitoring of digital works with the help of watermarks can be modified to the eHealth domain. The identities of clinicians who have viewed the EPHR can be embedded into the EPHR to maintain a log of access of the EPHR. The addition of a watermark to a cover work distorts the cover work. Addition of more watermarks increases the distortion. Hence the amount of data that could possibly be embedded into the EPHR is limited. The watermarks can be designed to contain a log of the last four/five health personnel who have viewed the EPHR.

- Fingerprinting: Fingerprinting is of two types: (i) the owner of the work would place a different watermark in each copy and record the recipient in each legal sale or distribution of the work. If the work is illegally distributed, the owner could find out who was responsible with the help of the watermark. (ii) a hash of the audio/visual work is created which is unique to the work. This fingerprint of the work helps track any manipulation and modification history within a signal without creating an overhead history file. The watermark records the list of transactions that have taken place in the history of the work in which the watermark is embedded.

  In an eHealth scenario, the fingerprinting of the EPHR can provide an application similar but not identical to that of fingerprinting multimedia content. One of the aims of digitising medical records is to provide researchers with access to biomedical data to further improve health in the community. Fingerprinting all the records belonging to different patients suffering from a similar illness, for example, skin cancer, with the same watermark can help researchers working on skin cancer to locate the required biomedical data easily. Second, embedding a unique fingerprint of every researcher who has accessed a huge set of data into the data will help trace the researcher responsible when large amounts of medical data are lost or distributed.

- Tamper proofing: Securely hide a signed summary of the work in a larger copy of itself. This can be used to prevent or to detect unauthorised modifications. Since the EPHR is used for purposes of diagnosis and research, any distortion to the cover could lead to misdiagnosis. Hence the distortion to the embedded messages must be minimum. Hiding a summary of the EPHR in the EPHR is not a viable solution. Instead by embedding a random string unique to the EPHR/hospital throughout the EPHR (in locations such that it distorts the EPHR to the minimum) a mechanism to identify any tampering of the EPHR could be achieved. The watermarks used for tamper proofing hence need to be semi-fragile/fragile.

- Provide different access levels to the data: By using different keys to embed the watermark, different access levels to the embedded data can be established. Since both embedding and decoding of the watermark depends on the key, knowledge of the key determines the amount of data that can be recovered at the receiver. For example, an administrator may be able to retrieve only the identification details of the patient while a clinician may be authorised to retrieve embedded information related to the medical condition of the patient.

1.2.2 A generic watermarking system

Figure 1.4 depicts a basic watermarking process. The message $\mathbf{m}$ to be embedded is generated based on the cover work $\mathbf{c}$ or independent of $\mathbf{c}$, to fulfil the different requirements of security and authentication. The derived message or watermark is usually transformed to a binary format WM.

$$\mathbf{m} \rightarrow WM. \tag{1.1}$$

Let $\mathcal{F}$ define the embedding technique used to embed the WM into $\mathbf{c}$. The locations of the samples of $\mathbf{c}$ used to carry the WM is represented by $\mathbf{k}$ and is referred to as the secret key. The length of $\mathbf{k}$ is dependent on the length of WM and $\mathcal{F}$. The watermarked cover $\tilde{\mathbf{c}}$ is obtained as

$$\mathcal{F}(\mathbf{c}(\mathbf{k}), WM) \rightarrow \tilde{\mathbf{c}}. \tag{1.2}$$

The watermarked document $\tilde{\mathbf{c}}$ is subjected to different common signal processing distortions η (assumed additive) during transmission resulting in an attacked watermarked

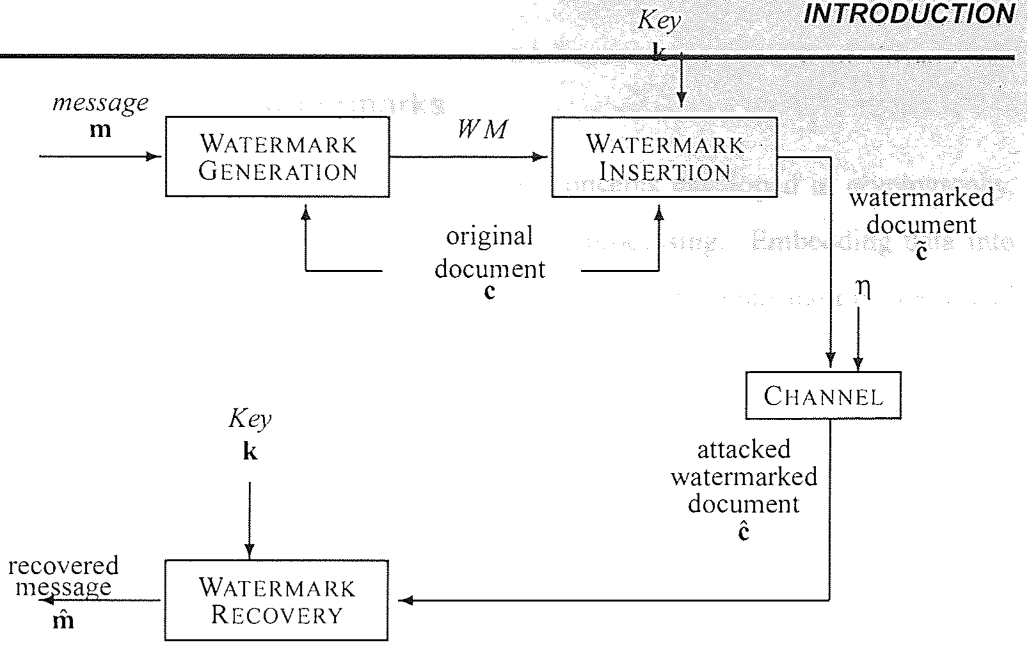


Figure 1.4: Block diagram of a generic watermarking system showing the process of watermark generation, embedding and transmission at the transmitter. The inverse technique of the watermark embedding and generation method is applied at the decoder to retrieve an estimate of the embedded watermark.

cover $\hat{\mathbf{c}}$,

$$\tilde{\mathbf{c}}+\eta \rightarrow \hat{\mathbf{c}}. \tag{1.3}$$

The inverse of the WM insertion and WM generation process is applied sequentially to $\hat{\mathbf{c}}$ at the decoder to obtain an estimate of the embedded WM, $\hat{WM}$ which in turn leads to the estimate of the embedded message $\hat{\mathbf{m}}$,

$$\mathcal{F}^{-1}(\hat{\mathbf{c}},\mathbf{k}) \rightarrow \hat{WM} \rightarrow \hat{\mathbf{m}}. \tag{1.4}$$

Applying prior knowledge of $\mathbf{m}$ even in the absence of $\mathbf{c}$, to $\hat{\mathbf{m}}$, authentication of the originality of $\hat{\mathbf{c}}$ and retrieval of the hidden message can be achieved. The Hamming distance between WM and $\hat{WM}$ when WM is known at the decoder provides a measure of the attack η. Hamming distance between two binary signals of equal length is defined as the sum of the bit positions where the two binary signals differ. This is obtained as the count of the number of ones in the XOR difference between the two signals.

$$dist(WM,\hat{WM}) = \sum(WM \oplus \hat{WM}). \tag{1.5}$$

In the absence of any knowledge of $\mathbf{m}$ at the decoder a predetermined threshold value can be used to determine the presence of $\mathbf{m}$.

1.2.3 Classification of watermarks

Watermarking techniques are based on ideas and concepts developed in cryptography, communication theory, algorithm design and signal processing. Embedding data into other digital data has been possible due to the limitations of the human auditory and visual systems (HA/VS) [45]. Watermarks are mainly classified based on the application of the watermark. They can be further classified on the embedding and decoding techniques used. One of the classifications based on the embedding method is shown in Figure 1.5 [84]:

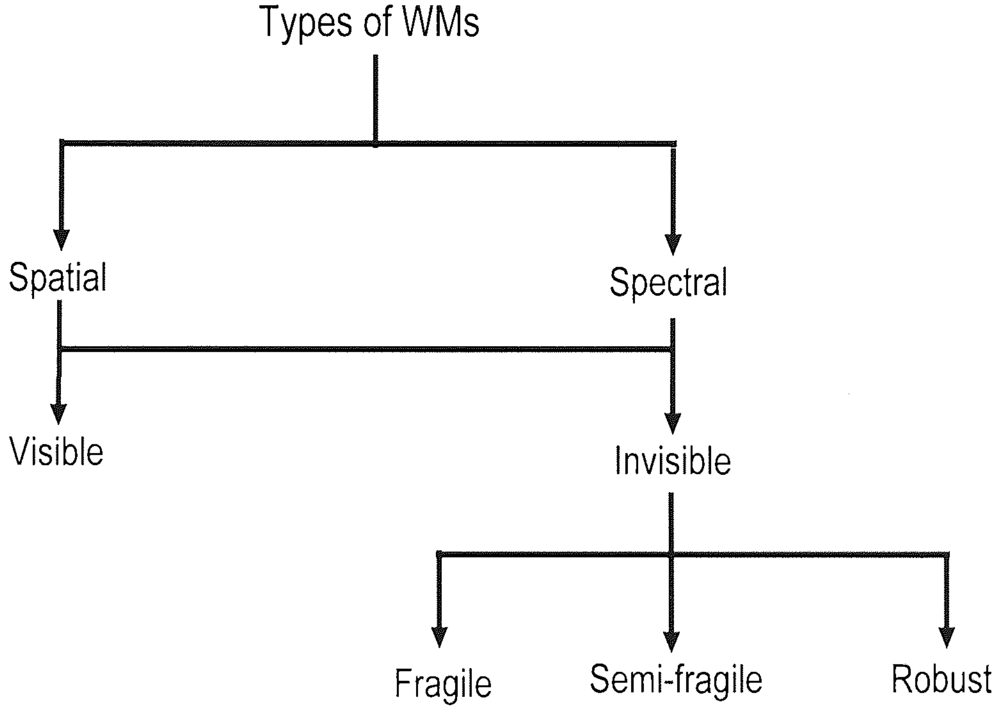


Figure 1.5: Classification of watermarks. This classification is based on the method of embedding and the application of the watermarks.

If the watermarks are embedded in the time domain representation of the host signal, the watermarks are referred to as spatial watermarks and if they are embedded in a transformed space of **c** the watermarks are known as spectral watermarks. Both spatial and spectral watermarks can be further classified as visible and invisible watermarks. Visible watermarks may be visual patterns (eg., a company logo or copyright sign) overlaid on digital images. They are designed to identify the owner and can be seen by every user of the data. The watermark cannot be removed from the original. Watermarks which are embedded in **c** and are perceptually transparent are called invisible watermarks. Invisible watermarks can be further classified as fragile - watermarks which are distorted due to slight alterations to $\tilde{\mathbf{c}}$, semi-fragile - watermarks which are distorted when η is large and exceeds a threshold value. The threshold is usually set by the owner of $\mathbf{c}$. Invisible watermarks which survive severe manipulation or tampering of $\tilde{\mathbf{c}}$ are known as robust watermarks.

The watermarks can also be classified as private or public (oblivious) when the decoding method requires the original or reference data for watermark detection or does not require the original data.

Since the watermarks used in the biomedical domain need to be secure and indistinguishable and robust to compression, they will usually be of the invisible type, robust and public. The watermarks used for authentication will also be invisible but fragile and public.

1.2.4 Characteristics of watermarks

The desired properties of every watermark vary with the application of the work (video, audio, images, text etc.,). These characteristics though defined by the application are limited by the embedding and decoding techniques, the communication system used, and the type of attacks they might face. Some fundamental characteristics of a watermark [77] are imperceptibility, rate of information, robustness and security.

Perceptual transparency:

The watermark must be embedded without affecting the perceptual quality of the underlying host signal. The procedure is imperceptible if the Human Audio/Visual System (HA/VS) cannot differentiate between the original host signal and a host signal with inserted data. Imperceptibility is usually determined by blind testing of the watermarked data. One such example test procedure used to measure perceptual phenomena is the two alternative, forced choice (2AFC) [36]. Observers are randomly presented with signals with and without embedded data and asked to determine the higher quality signal. If 50% of the observers correctly identify the watermarked content the imperceptibility is termed as zero just noticeable difference (JND). JND is a measurement unit used for psychophysics studies and it represents a level of distortion that can be perceived in 50% of the experimental trials. If 75% of the observers correctly identify the watermarked content then the watemarked content is said to have one JND and it is indicative of the existence of the WM. The main criterion of data embedding is that it should not produce perceptually dissimilar artifacts. It must also take care of typical modifications that the signal may undergo. e.g., digital pictures typically undergo sharpening or high pass filtering. The distortion to the cover work due to the embedded watermark $\mathcal{D}_{Emb}$ is given by

$$\mathcal{D}_{Emb} = \tilde{\mathbf{c}}_i - \mathbf{c}_i. \tag{1.6}$$

$\tilde{\mathbf{c}}_i$ and $\mathbf{c}_i$ represent the individual elements of the watermarked and original work respectively with $1 \leq i \leq \mathbf{Nsamp}$ where **Nsamp** is the total number of samples/ pixels of the cover work. Based on the application, an acceptable level of $\mathcal{D}_{Emb}$, ϑ is decided upon at the watermark embedder. Imperceptibility is therefore not defined by the value of $\mathcal{D}_{Emb}$ but by the maximum value that it can assume for a given application.

$$\mathcal{D}_{Emb} \leq \vartheta. \tag{1.7}$$

Evaluation of imperceptibility of a WM:
Perceptibility is a characteristic based on the HVS and is applicable to images or any form of data that can be visualised. The HVS [28] has a limited sensitivity, it does not react to small stimuli and is not able to discriminate between signals with an infinite precision. It also presents saturation effects. Since watermarking is the capability of the host signal to hide another signal, it can be termed as masking. Different masking phenomena exist for different stages of the HVS. The most common types of masking used in image processing include spatial masking (edges in images can mask signals of much greater amplitude than region of near-constant intensity) and contrast or pattern masking.

Image watermarking is based on the concepts studied in image processing. The ability of the cover image to hide the secret message is measured in terms of the visibility of the message. These perceptibility measurement terms are explained as applied to images, but they can also be used with time series data such as EEG and ECG, as the clinician views the ECG and EEG to make a diagnosis.

A list of selected measures to characterise watermarked systems are listed [79]:

- Relative entropy or the Kullback-Leibler divergence:

  Relative entropy, or the Kullback-Leibler divergence $\mathcal{KL}$, normalises the entropy of a watermarked signal $\tilde{\mathbf{c}}$, with respect to a reference signal $\mathbf{c}$.

$$\mathcal{KL} = \sum_i \mathrm{p}_i \log\left(\frac{\mathrm{p}_i}{\mathrm{q}_i}\right) \tag{1.8}$$

where p and q are the probability distributions of $\tilde{\mathbf{c}}$ and $\mathbf{c}$ respectively, over all sample values i. This is a measure of dissimilarity and used to represent mutual information. It is zero iff the two probability densities are zero. The larger the value of the mutual information, the more similar the two signals are to each other and vice versa.

- Peak Signal-to-Noise Ratio (PSNR):

$$PSNR = 10\log_{10}\left(\frac{B}{\mathbf{rms}}\right) \tag{1.9}$$

where B is the largest possible value of the signal or the bandwidth and **rms** is the root mean square difference between the two signals. If the value of $PSNR$ is large it indicates that the noise or distortion due to the embedded WM is very small.

- Mean Square Error (MSE):
  Compares two signals on a sample by sample basis.

$$MSE = \frac{1}{\mathbf{Nsamp}} \sum_i ||\mathbf{c}_i - \tilde{\mathbf{c}}_i||^2 \tag{1.10}$$

where both the signals, the reference signal $\mathbf{c}_i$ and actual image $\tilde{\mathbf{c}}_i$ contain **Nsamp** samples. For the WM to be imperceptible the MSE or difference between the two signals must be relatively small. There are obvious disadvantages in using MSE unless relative phase is known exactly.

Rate of information of the embedding algorithm:

The rate of information $\mathcal{R}$ is defined as the ratio of the length of a watermark to the length of $\mathbf{c}$. $\mathcal{R}$ typically depends on the application of $\mathbf{c}$ and the embedded watermark. Fraud detection applications require small amounts of information (insertion of serial number, author identification) incorporated repeatedly into the host signal, but embedding a smaller image into a larger image or multiple speech signals into a video requires a lot of bandwidth [93]. The data embedded can be a significant portion of the data in the host signal. Therefore the amount of data embedded depends critically on the embedding algorithm, the underlying host signal and most importantly the application of the watermark.

$$\mathcal{R} = \frac{|WM|(bits)}{|\mathbf{c}|(bits)}. \tag{1.11}$$

Robustness:

This is the ability of the embedded watermark to withstand distortions to $\tilde{\mathbf{c}}$. Unintentional distortions are due to common signal processing operations. Most applications use lossy coding operations to reduce bit rates and increase efficiency during storage and transmission. Also digital data can be easily modified and manipulated. A damaged host signal results in damaged embedded data. Malicious or purposeful modification/removal of the host signal is done to thwart the detection of the embedded data. Additional watermarks may also be embedded by malicious attackers to cause ambiguity in copyright applications. Hence watermarks that can be retrieved after common signal processing operations and malicious attacks are said to have a high level of robustness.

Evaluation of robustness of a watermark: Robustness of a watermark defines the capacity of a watermark to remain unchanged under any form of intentional/unintentional change to the watermarked cover. The various levels of robustness based on the type of attack are listed as follows [77]:

- Level zero: No special robustness features have been added except the ones needed to fulfil the purpose and operational environment of the scheme.

- Low level: Robustness features added to the watermarking technique but which can be circumvented using simple and cheap tools available publicly. These features are added to prevent "honest" people from disabling the mark during normal use of the work.

- Moderate robustness: Expensive tools are required as well as some basic knowledge on watermarking to disable the watermark.

- Moderately high: Tools are available but special skills and knowledge are required and attempts to distort the watermark may be unsuccessful. Several attempts and operations may be required and one may have to work on the approach.

- High robustness: All known attempts have been unsuccessful. Some research by a team of specialists is necessary. The cost of the attempt may be much higher than what it is worth and its success is uncertain.
- Provable robustness: It means it should be computationally (or even more stringent: theoretically) infeasible for a willful opponent to disable the watermark.

The multiple watermarks denoting the personal information and added information, embedded into the EPHR are characterised based on the level of robustness required by each watermark. This will be discussed in more detail in chapter 4.

Security:

The embedding procedure must be secure in that an unauthorised user must not be able to detect the presence of embedded data nor remove the embedded data [50]. A data embedding procedure is said to be secure if knowing the exact algorithm for embedding the data, does not help an unauthorised party to detect the presence of the embedded data. This is because the unauthorised user does not have access to the secret key that controls the insertion of the data in the host signal.

The information that needs to be embedded can be further secured by first encrypting the information and then embedding into the cover. A detailed discussion of this approach will be presented in chapter 3, section 3.2 (This work was presented in [56, 57].

Some additional desired properties are fast information embedding and/or retrieval, compressed domain processing, statistical undetectability.

The first three characteristics of imperceptibility, capacity and robustness are trade-offs against each other and can be represented as a triangle as shown in Figure 1.6. The trade-off is obtained based on the requirements of the problem domain.

Since biomedical data is mainly used for diagnosis, the imperceptibility of the watermark should be as high as possible. Distortions to the original due to the watermark may result in wrong interpretation of the data. Also malicious attacks to distort or completely remove the watermark are not expected. If the embedded data is robust to simple signal processing techniques necessary for efficient transmission, it is sufficient [79].

Hence in figure 1.6, the preferred location of the Biomedical Watermarking in the trade-off triangle is as shown, but this is not a fixed position. Some specialist applications

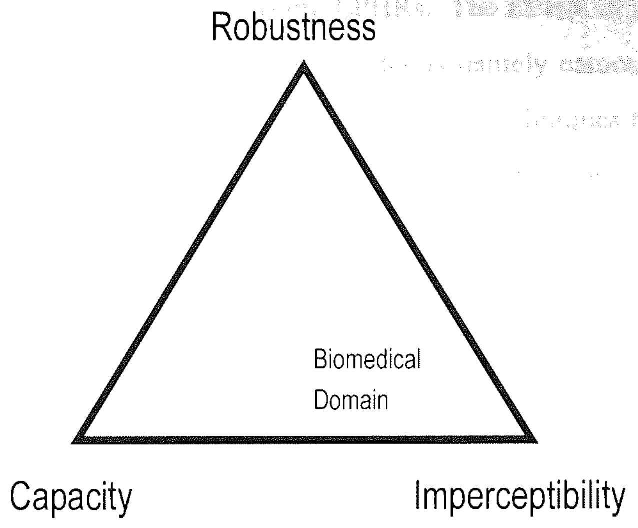


Figure 1.6: Characteristics of a watermark; Trade-off Triangle

of the medical data might require a different level of robustness, in which case this position might vary. The methods considered in this thesis need to be flexible enough to meet these criteria in the trade-off triangle.

1.2.5 Watermark embedding techniques

Watermark embedding techniques are of two types, blind and informed [65]. The classification here is based on the method of generating the message. In the blind embedding technique, the watermark is generated independently of the cover work in which the watermark is to be hidden. The watermark message is chosen from a set of messages based on a key **k**. This type of watermark may cause huge distortions to the cover work and have a low level of imperceptibility. Hence appropriate scaling of the watermark is required to make it imperceptible. In the informed embedding technique, the watermark is chosen based on the type of cover work. Hence the watermark does not distort the original significantly.

The different embedding techniques are distinguished based on the information rate, selection of the location in the cover where the message is embedded and types of messages embedded [78]. Since the EPHR cannot be altered to suit the embedded message, the embedding of watermarks cannot be based on the principle of security through obscurity. Also the EPHR is mainly used for diagnostic purposes and as already mentioned, the value of $\mathcal{D}_{Emb}$ should be minimum. Spreading the hidden information, increases $\mathcal{D}_{Emb}$, hence it is not applicable to watermarking EPHRs. The EPHR can be watermarked using a combination of the other embedding methods namely camouflage and hiding the location of the hidden information [78]. Watermarking techniques that enable the use of automated verification would also be useful in the authentication of the EPHR. The unwatermarked cover or the embedded information cannot be transmitted to the watermark detector as the security of the embedded information will be compromised. The work presented in [44] proposed a method for the verification of the received data by using a local average scheme. In this method the image data is segmented into blocks and an average of each block and the size of each block are transmitted. The watermark detector divides the received image into similar blocks and determines the average of each block. The average values obtained by the watermark detector are compared with the average values received from the watermark embedder. The differences obtained as a result of the comparison are used to detect minor changes and also localise the distortions to the waterked cover during transmission.

1.2.6 Attacks on watermarking schemes

The types of attacks on watermarking schemes are based on the level of distortion caused to the watermark and the intent of the attacker [93, 26]. These can be classified as:

- Robustness attacks:
  Robustness attacks are malicious attacks which attempt to completely destroy or remove the watermark without considerable damage to the original data. This type of attack may occur when the embedded watermarks are used as a log of access of the EPHR.

- Presentation attacks:
  The watermark is changed in such a way that the detector will not be able to find it. Complete removal of the watermark is not the aim of these attacks. This class of attacks is similar to robustness attacks and in an eHealth scenario these attacks may be utilised by an unauthorised accessor of the EPHR to delete any record of their access of the EPHR.

- Collusion attacks:
  In this type of attack, the attacker obtains multiple copies of the cover wherein each

cover is embedded with a different watermark. The attacker combines the various covers with different watermarks and constructs a single cover with no watermark.

- Interpretation attacks:

  These attacks seek to confuse the decoder by embedding different watermarks in an already watermarked cover. This leads to ambiguous decisions and deadlock in case of ownership disputes. Interpretation attacks in an eHealth scenario could occur when an attacker requires another individual's EPHR to lay claim to certain benefits for example buying insurance, applying for certain jobs.

- Legal attacks:

  Legal attacks take advantage of existing copyright laws to create doubts on watermarking schemes. This form of attack is usually seen in copyright disputes of work. The possibility of legal attacks in an eHealth scenario is minimal.

Since the embedded watermark is typically expected to be the personal information of the patient to whom the EPHR belongs, obtaining the exact embedded message will be the aim of the attacker. Without the knowledge of the personal information, every EPHR is similar to every other EPHR in terms of usefulness to the attacker. Watermark embedding must therefore be robust to most signal processing techniques that might be applied to an EPHR in order to store it in a compact format and still render it meaningful for diagnostic purposes. Hence ensuring that it is robust to most of the modifications mentioned above is necessary.

1.2.7 Watermark extraction

Watermark extraction is of two types, blind and informed [24]. The classification is defined on the availability of the original cover at the detector. In the blind form of extraction, the original cover is not available for watermark detection as in the case of the EPHR, whereas in the informed extraction process the original is available for extraction. In multimedia applications such as copy tracking, copyright protection etc., data extraction algorithms have the provision to use the original signal. When the received signal can be compared with the original to obtain the embedded data, any distortions to the watermarked signal can be easily determined. Hence techniques that use the original signal are robust to a larger assortment of distortions.

But in most other applications as in the biomedical domain there is no access to the original unwatermarked signal while extracting embedded data. As already mentioned, in an EPHR the value of $\mathcal{D}_{Emb}$ should be minimum. This criterion limits the data that can be embedded, and as the watermark extraction is blind it makes the extraction of the embedded message difficult. This problem is intensified when the embedded data or information on the communication channel is corrupted by strong interference and channel effects.

1.2.8 Evaluation of watermarking techniques

Watermarking techniques are evaluated on the reliability of the decoder in the presence of attacks, ie the number of false positives, and false negatives. False positive is when the decoder detects a watermark in the received data in the absence of any embedded data and false negative is when the decoder cannot detect a watermark in the received data though a watermark is present. A high rate of false positive/false negative indicates that the performance of the watermarking technique is very low.

Watermarking applications in the biomedical domain require the detector to not only detect the presence of the watermark but also decode it, in the presence of noise (additional embedded watermarks, common signal processing techniques such as compression, scaling, A/D, D/A conversions, malicious attacks). Estimation of the correctness of the retrieved watermark can be assessed by different methods, some of which are mentioned below.

- Linear correlation(LC): Correlation of the estimated watermark, $\widehat{WM}$ with the original watermark, WM either in informed or blind detection is the most commonly used decoding method in most of the watermarking applications [24]. $\widehat{WM}$ is correlated bit by bit with the original message WM and the resulting Hamming distance $dist$ is compared to a threshold ρ.

  The output of the decoder, say y, is as follows:

$$y = \begin{cases} 1, & \text{if } dist > \rho \\ 0, & \text{if } dist \leq \rho, \end{cases} \tag{1.12}$$

  where a '1' indicates the presence of a WM and a '0' an absence of WM.

- Similarity metric: This evaluation method was described by Cox et.al [25] to establish that a false positive judgement is highly unlikely. The authors based their argument on the rare chance of $\hat{\mathbf{m}}$ being identical to the original $\mathbf{m}$. The similarity between $\hat{\mathbf{m}}$ and $\mathbf{m}$ is measured as

$$Sim(m,\hat{m}) = \left(\frac{m.\hat{m}}{\hat{m}.\hat{m}}\right). \tag{1.13}$$

If an attacker distorts the watermarked cover $\tilde{\mathbf{c}}$ to produce $\hat{\mathbf{c}}$ without any access to the original unwatermarked data $\mathbf{c}$, then even with a fixed value of $\hat{\mathbf{m}}$, each sample $\tilde{\mathbf{c}}(\mathbf{k})$ representing $\hat{\mathbf{m}}$ will be independently distributed according to $\mathcal{N}(0,1)$. Hence $sim(\mathbf{m},\hat{\mathbf{m}})$ will also be distributed according to $\mathcal{N}(0,1)$. Then according to standard significance tests for the normal distribution, it is extremely unlikely that $sim(\mathbf{m},\hat{\mathbf{m}}) > 6$ on a scale of 1 to 10.

Though the original embedded message is not available at the decoder in the biomedical domain since the WM refers to the personal information in the EPHR and is known only at the embedder, the LC method of evaluating the watermarking technique and robustness of the embedded WM will be used. This is because it is the simplest method to test the design of a watermarking system which is suitable for the biomedical domain.

1.3 Use of Watermarking Principles in an eHealth System

In this chapter a discussion of the issues involving EPHRs and reasons why they cannot be resolved with the use of smart cards or role-based access networks only was presented. The advantages and properties of watermarks (both steganographic and non-steganographic) that help design a mechanism to secure patient privacy is also given, since embedding patient details in a bioprofile has the advantage that the embedded data is imperceptible and robust to attacks. The embedded data cannot be removed or distorted separately from the cover [24]. Also extra security can be provided to the data by encrypting it before embedding it.

Since the patient's details are embedded in the bioprofile, the bandwidth of the information transmitted is less than the information in which the patient's details are encrypted. When data is encrypted it is extra to the actual medical recordings as it is attached as a header to the medical record. If instead this header is embedded inside the medical record, the data that is actually transmitted is the medical recording only. In the medical domain where quality of service (quick access to the EPHR) impacts on life-critical decisions, real-time access and continuous availability of bulk data (EPHR) are required. Large waiting times thus can be critical to a patient's life.

1.4 Summary and Thesis Overview

In this Introduction the requirements of privacy in the EPHR, and basic properties of information hiding principles and advantages were surveyed. The conclusion arrived at is that, embedding EPHRs with the personal information in the medical record is the best possible solution available currently to protecting patient privacy. This thesis discusses the design of a new data embedding technique capable of being implemented in a secure eHealth network (chapter 2). It will be shown how the data hiding approach has an inbuilt defence against data snooping (chapter 3). In addition, it will be revealed how the framework can be extended to multiple watermarks of similar/dissimilar characteristics. The multiple watermarks will be embedded into a one-dimensional cover to demonstrate the capability of the new method (chapters 4, 5). In chapter 6 a mathematical discussion of the security of the new data hiding approach is presented.

The thesis focusses on one-dimensional biomedical data. This is because the low dimensionality and reduced redundancy poses greater challenges to watermarking than the typical image processing domain usually considered. However, the approach in this thesis can be extended to higher dimensional biomedical signals. The thesis is the work of the author but parts of it have appeared in the public domain. A list of publications has been presented at the end of the thesis.

2 WATERMARKING SINGLE CHANNEL TIME-SERIES DATA

CONTENTS

The issues of patient privacy and security of personal information in medical records and the lack of correct security protocols in EPHR standards that are currently being implemented and tested were discussed in the previous chapter. The applications of watermarking for multimedia data and their relative success discussed in the last chapter indicated that watermarking of EPHRs could help resolve the security issues of EPHRs. The potential advantages watermarking methods can provide complementary to cryptographic and ICT based security protocols was discussed. In this chapter the development of watermarking in the context of the EPHR is implemented.

Some commonly used watermarking techniques for medical data have been demonstrated in the following transform domains: Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) [70], Discrete Wavelet Transform (DWT) [33], Principal Component Analysis (PCA) [96]. Other preliminary research into the securing of privacy and prevention of tampering, of medical records by watermarking them has been presented in [6, 100, 27, 19, 80, 80, 64, 9, 37, 102].

Preliminary research into watermarking biomedical data has mainly concentrated on biomedical images since a large portion of medical data is in the form of images. All of the above references except the work by Toch et al [96] is based on medical image data. However, techniques developed for images do not transpose well to other data modalities. Embedding data in a single channel time series which has a low redundancy is much more difficult due to the reduced redundancy limiting possibilities of hiding data and has not been investigated. Time series medical data such as EEG and ECG are important in helping diagnose a large number of health problems and conditions. Identifying an EPHR standard that performs equally for any input data format (image, time series) is necessary. Secondly, different messages including patient identity, doctors' notes, record of clinicians who have accessed the medical record require different levels of security. A watermarking system capable of identifying different robustness zones to embed the different watermarks hierarchically is needed.

Hence, the aim of this chapter is to consider appropriate efficient watermarking algorithms suitable for single channel time series data and also capable of providing signals with differing levels of robustness for embedding multiple watermarks. The signal processing aspects of a frame-based approach of expanding signals using a nonorthogonal basis derived from the data is exploited, and the consequences of applying the framework to different dimensionality biopatterns is investigated. A study of transform domain methods as opposed to time domain methods is conducted, as obtaining a hierarchical system such as mentioned above is not possible in the time domain of the data. The time domain of the data gives a continuous description of the data for different instances of time. It does not define regions of the data that are robust to specific attacks such as filtering and compression. Transforming the given data from the spatial domain to a frequency domain helps identify specific regions of the transform domain which are robust against different levels of filtering and compression. Peining et al [94] also observe that for image watermarking systems 'in general, the systems that embed the watermark in the pixel domain are less robust to image manipulations, and semi-blind and blind systems are more prone to false positives (detecting the watermark in an unmarked image) and false negatives (not detecting the watermark in a marked image)'. This is true of all watermarking systems including time series watermarking systems.

The integrity of the hidden message for different attacks such as sampling rate, low pass filtering, addition of Gaussian noise and compression is investigated. Finding a transform suitable for designing a hierarchical multiple watermarking system robust under compression is the main motivation of this chapter.

2.1 Transform Domain Watermarking Methods

Embedding watermarks in $\mathbf{c}$ requires replication of the watermarks and spreading of the watermarks to cover the entire length of $\mathbf{c}$ applying spread spectrum based techniques [25]. The watermark signal is of low intensity compared to the strength of $\mathbf{c}$. This is to ensure that the value of $\mathcal{D}_{Emb}$ is minimum and the watermark is robust against various signal processing attacks, mainly compression. Orthogonal transforms such as the DFT, DCT, DWT and PCA instead decorrelate the components of $\mathbf{c}$ and redistribute the energy of $\mathbf{c}$ such that it is contained in a few components. Compression of a given data is achieved by first transforming the data using one of the above mentioned transforms. The original signal is reconstructed using only those components of the transformed representation containing the maximum energy. By embedding the watermarks in these components containing a significant amount of energy of $\mathbf{c}$ thus ensures that the watermarks are robust to compression. Though many different transform domain techniques exist and have been studied, five transforms which provide a good transformation into the spectral domain of $\mathbf{c}$ or translation into different representations of $\mathbf{c}$ will be further investigated in this thesis. The transforms studied are DFT, DCT, DWT, PCA and ICA. DFT and DCT give the spectral representation of **c** while DWT, PCA and ICA transform the input **c** into projections onto basis vectors. Each method derives the basis vectors differently, hence the projections obtained for each method vary, have different characteristics and can be applied differently.

2.1.1 Discrete Fourier transform

The Fourier transform of a continuous time, aperiodic signal gives the frequency analysis of the signal [82]. The Fourier transform of **c** represented over time t is given by $C(F)$.

$$C(F) = \int_{-\infty}^{\infty} \mathbf{c}(t) e^{-j2\pi F t} dt. \tag{2.1}$$

Identifying the frequency spectrum of a noisy signal which is a composite of multiple

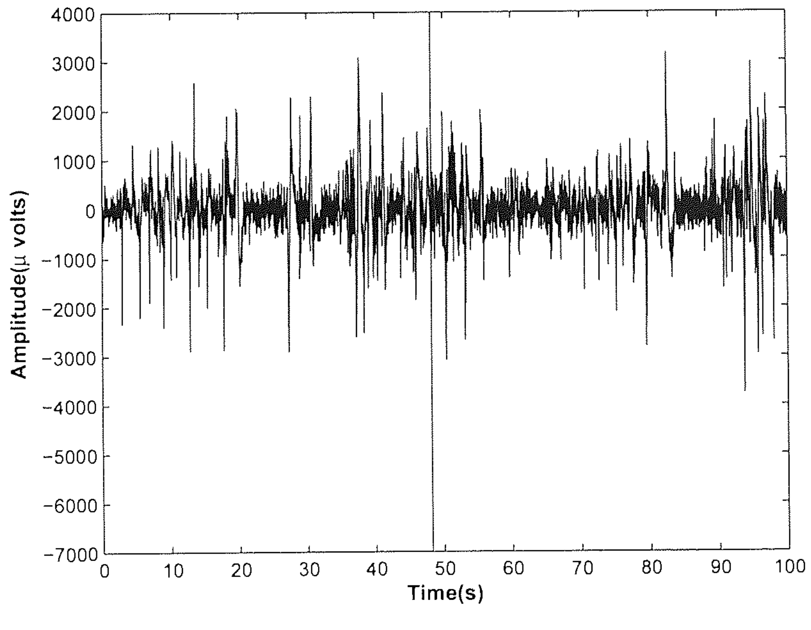


Figure 2.1: Time domain representation of the EEG.

signals of differing frequencies is not possible by studying the time domain representation of the signal (see figure 2.1).

The spectrogram showing the amplitude of a given frequency calculated over a period of time is shown in figure 2.2. This figure illustrates the complexity of the EEG signal.

Filtering attacks usually filter out the high frequency noise to obtain a better representation of the data under observation. Hence embedding information in the low and middle

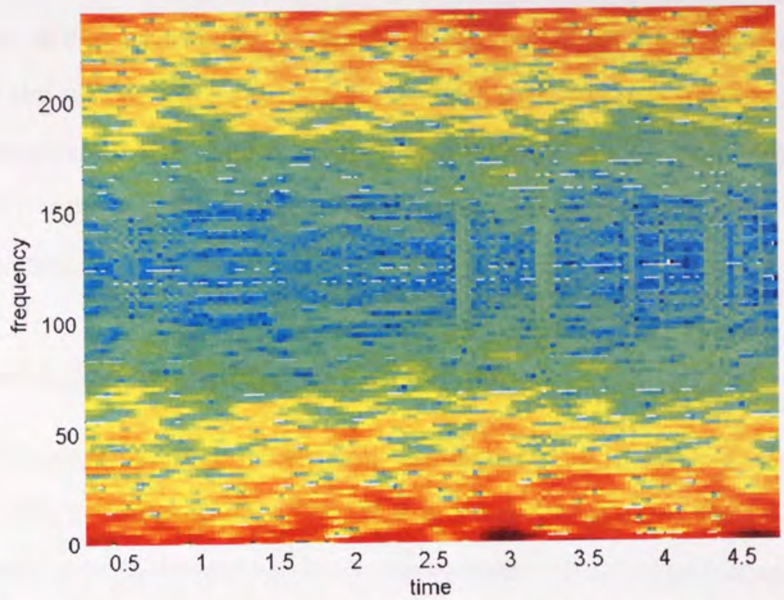


Figure 2.2: Spectrogram of the EEG denoting the time-frequency representation of the EEG.

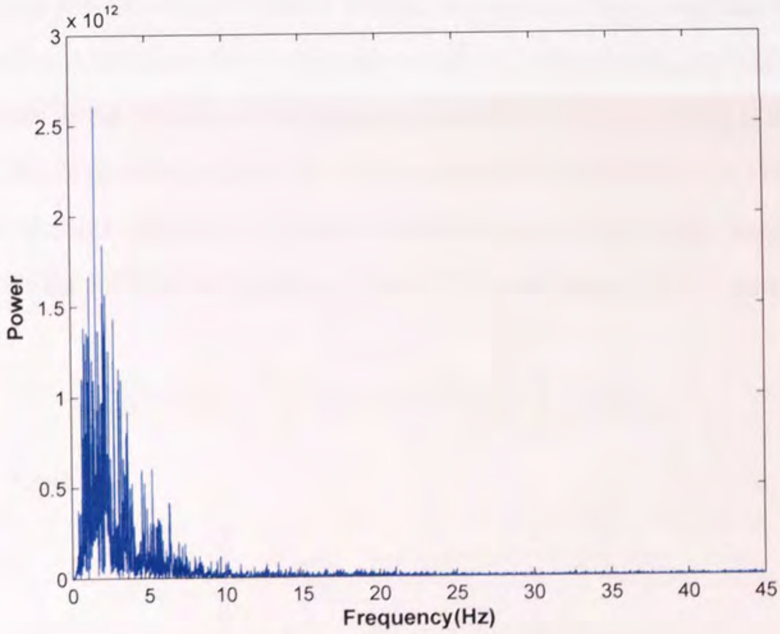


Figure 2.3: Spectrum of the EEG.

frequencies of a signal is a better strategy to preserve the embedded message after filtering. But embedding information in the low frequency regions of **c** increases the value of $\mathcal{D}_{Emb}$. This is because the low frequency components of the signal contain most of the information of the signal and hence any distortion has a larger effect here. Instead embedding information in the middle frequency representations of **c** reduces the distortion, $\mathcal{D}_{Emb}$ and also survives filtering attacks.

Figure 2.3 shows the long term power spectrum of $\mathbf{c}$, $|C(F)|^2$. This figure gives the power content of $\mathbf{c}$ at a particular frequency. A random selection of samples $\mathbf{k}$ equal to the length of the watermark are chosen from $C(F)$ such that they represent the middle frequency components of $\mathbf{c}$. The watermark is embedded in $C(F)(\mathbf{k})$ samples using $\mathcal{F}$ to obtain $\tilde{C}(F)$. Applying the inverse of the Fourier transform to $\tilde{C}(F)$, the watermarked cover $\tilde{\mathbf{c}}$ is obtained.

2.1.2 Discrete cosine transform

The discrete cosine transform converts the given input signal in terms of a sum of cosine functions of different frequencies. The DCT is a specialisation of the DFT and is the most commonly used technique for lossy compression of audio and images. Most of the watermarking and compression methods were mainly developed for images. Image compression is obtained by dividing the image into blocks and compressing each block. The DCT produces less blocking effects in the decompressed image compared to the DFT. Where compression is obtained over finite sections of an infinite signal, the DCT produces less discontinuities at the boundaries compared to the DFT. When a signal is transformed using the DCT, the original signal can be effectively reconstructed using a small number of its DCT co-efficients. Hence DCT based watermarking methods are studied though they are similar to the DFT based methods. The DCT co-efficients of $\mathbf{c}(t)$ (see figure 2.4) are derived as

$$C(u) = \alpha(u) \sum_{t=1}^{\mathbf{N}_{samp}} \mathbf{c}(t) cos(\frac{\pi(2t-1)(u-1)}{2\mathbf{N}_{samp}}) \tag{2.2}$$

where u = 1 to $\mathbf{N}_{samp}$ and

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{\mathbf{N}_{samp}}} & \text{for u = 1} \\ \sqrt{\frac{2}{\mathbf{N}_{samp}}} & \text{for } 2 \leq \text{u} \leq \mathbf{N}_{samp} \end{cases} \tag{2.3}$$

$\mathbf{N}_{samp}$ is of the length of $\mathbf{c}$ and, $\mathbf{c}$ and C are the same length.

The magnitude of the cosine function gives the measure of information content of $\mathbf{c}$ represented by the cosine transform. A large magnitude value of $C(u)$ represents the low frequency component of the signal and hence contains a large amount of information of $\mathbf{c}$.

The cosine transform being similar to the DFT (gives a spectral representation of the signal, see figure 2.4), modifying the samples of C of large magnitude ensures that the

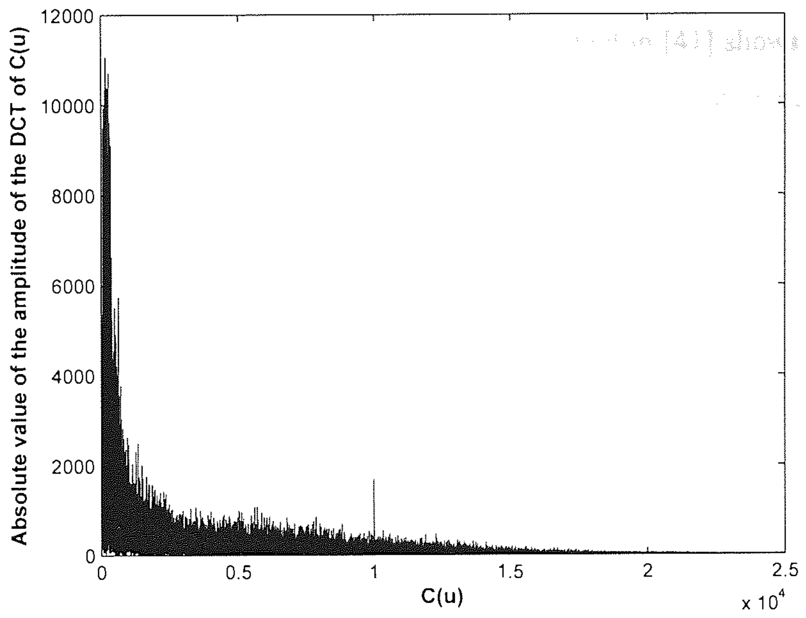


Figure 2.4: DCT values of the EEG.

embedded watermark is robust to compression but the distortion to **c** is higher. Hence **k** denotes the samples of C representing the middle frequency components of **c**. Applying the inverse of the cosine transform to $\tilde{C}(u)$, the watermarked cover $\tilde{\mathbf{c}}$ is obtained.

2.1.3 Discrete wavelet transform

The DFT and DCT transforms provide representations of the amplitude of the frequencies in **c**. In certain applications obtaining a time-frequency plot of the signal is advantageous. The spectrogram based short-time Fourier transform (figure 2.2) and DWT are used to obtain time-frequency analyses of a signal. While the spectrogram provides information about all the frequencies for different periods of time, the DWT provides a good time-frequency localisation of the data by scaling the length of, and shifting the mother wavelet over the signal under consideration [71]. The wavelets' capability of providing good time resolution at high frequencies and good frequency resolution at low frequencies has made it a popular tool for medical signal processing [95] and compression algorithms. Embedding a digital watermark in the wavelet transform domain of digital images has been presented in various works. Listing all the publications and their contributions is beyond the scope of this thesis. Hence a sample of works presented on watermark embedding in multilevel wavelet decomposition [41, 98] of the host data and their contributions are mentioned. [98] discuss a watermark embedding method based on the wavelet transform which is resilient to collusion attacks. The work presented in [41] shows how an image representing a watermark can be embedded into a host image, by utilising the properties of the wavelet co-efficients obtained for multilevel wavelet decomposition of the host image. They show that the watermark embedding method is robust to a variety of signal processing distoritions (including JPEG, image cropping, sharpening, median filtering) of the watermarked host image.

Each wavelet decomposition of the original signal halves the frequency and length of the signal. The Haar wavelet function is defined as

$$\psi(t) = \begin{cases} 1 & \text{for } 0 \leq \mathrm{t} < 1/2, \\ -1 & \text{for } 1/2 < \mathrm{t} \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The filter co-efficients for the Haar function are shown in figure 2.5.

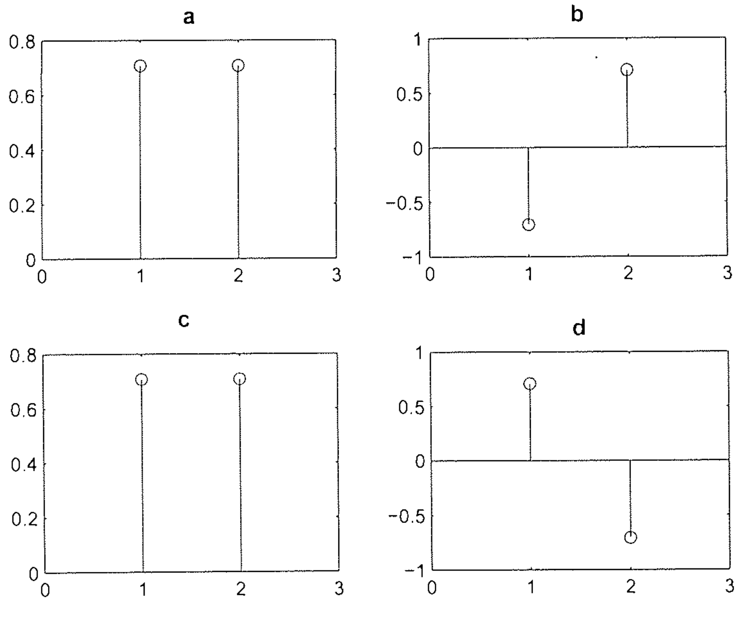


Figure 2.5: The decomposition and reconstruction filters' co-efficients of the Haar wavelet $\psi(t)$ defined above. a.Decomposition co-efficients of low pass filter b.Decomposition co-efficients of high pass filter c.Reconstruction co-efficients of low pass filter d.Reconstruction co-efficients of high pass filter.

The Haar function is a common choice to generate wavelets from time series data [74]. For time series data with continuous random changes the Haar wavelet is more suitable for fast changing time-series such as an EEG compared to Daubechies wavelets, Mexican

Hat wavelets and Morlet wavelets. The Daubechies, Mexican Hat and Morlet wavelet algorithms are better suited for smoothly changing time series. The Haar wavelet is also simple, fast and exactly reversible.

The Haar function ψ used as the mother wavelet generates a set of wavelets

$$C_{a,b} = \sum_{\mathbf{N}_{samp}} \mathbf{c}(t)\psi_{a,b}(\mathbf{t}). \tag{2.5}$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{\mathrm{s}}}\psi(\frac{t-\tau}{\mathrm{s}}), \tag{2.6}$$

where a denotes the dilation index, b the translation index, s the scale factor and τ the displacement, and a and b are integers. The DWT is basically the application of a set of filters (figure 2.5 (a) and (b)) to $\mathbf{c}$ (figure 2.1) resulting in an approximate C_a, and fine detailed C_b representation of $\mathbf{c}$.

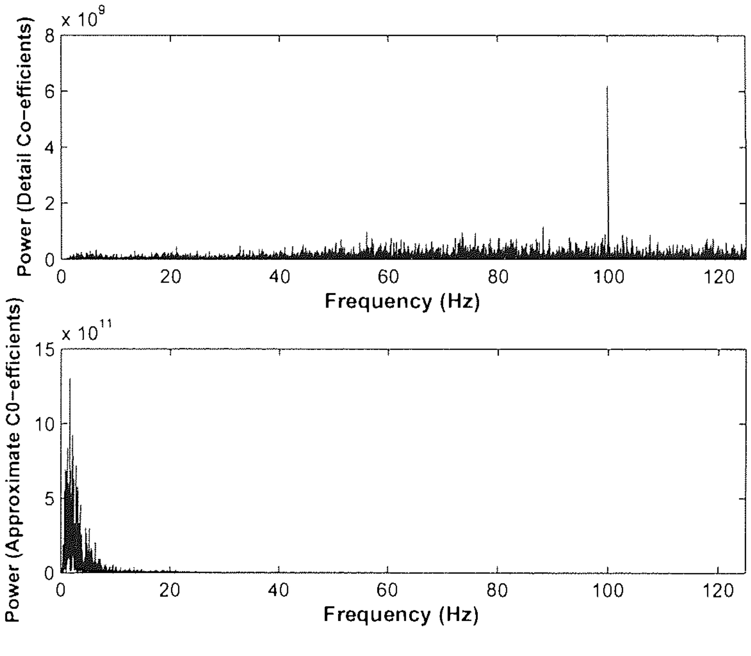


Figure 2.6: Frequency content of the approximate and detail co-efficients.

Let $C_{a_{init}}$ denote the resulting signal obtained by the application of the filter shown in figure 2.6 (a) to $\mathbf{c}$. The filter is applied to each pair of samples of $\mathbf{c}$, $[\mathbf{c}_i, \mathbf{c}_{i+1}]$ where $i \in [1, \mathbf{N}_{samp}]$. C_a is the signal obtained by the downsampling of $C_{a_{init}}$ by a factor of 2. Similarly, let $C_{b_{init}}$ denote the resulting signal obtained by the application of the filter shown in figure 2.6 (b) to $\mathbf{c}$. As in the case of C_a, the filter is applied to each pair of samples of $\mathbf{c}$. Alternate samples of $C_{b_{init}}$ are used to obtain the signal C_b.

Figure 2.6 gives the frequency spectrum of C_a (top) and the frequency spectrum of C_b (below). These spectra are obtained by applying the Fourier transform to C_a and C_b respectively. The plots of figure 2.6 give an indication of the spectra of C_a and C_b but do not identify the exact samples of C_a and C_b which correspond to a particular frequency range. Hence in order to maintain a low distortion rate, **k** in the DWT method represents a selection of samples of the detail co-efficients. Though the detail co-efficients are basically noise components, the main restriction for watermarking biomedical data, is to maintain a very low distortion to **c**, hence the detail co-efficients and not the approximate co-efficients are used to embed the watermark. The middle frequency components of the signal can be obtained by successive application of the DWT transform which will be explored in the next chapter. Applying the inverse of the wavelet transform (filters shown in figure 2.5 (c) and (d))to $\tilde{C_b}$ and C_a, the watermarked cover $\tilde{\mathbf{c}}$ is obtained.

2.1.4 Principal component analysis

PCA is one of the most commonly used techniques in statistical data analysis, feature extraction and data compression applications [42]. It is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system. The first principal component gives the projection of the data with the largest variance. The second greatest variance on the second coordinate, and so on. PCA is theoretically the optimum transform for a given data in least square terms. A high dimensional data set is transformed into a low dimensional data set which can be easily visualised by only retaining a few components. PCA utilises the first and second order statistics of the data. The redundancy of each variable of the input data is measured by calculating the covariance matrix of the input data.

PCA of a given data set is based on the following assumption:

- Assumption on Linearity.
- Assumption on the statistical importance of mean and covariance.

Given a one-dimensional observation **c** of length $\mathbf{N}_{samp}$, the data matrix $\mathbf{X}_{p \times N_{nov}}$ is constructed as follows. Let each row of $\mathbf{X}$, $\mathbf{x}_i$ represent an observation of length N_{nov}.

$$\mathbf{x}_i = \mathbf{c}[(i * N_{nov}) + 1, \ldots, (i-1) * N_{nov}]. \tag{2.7}$$

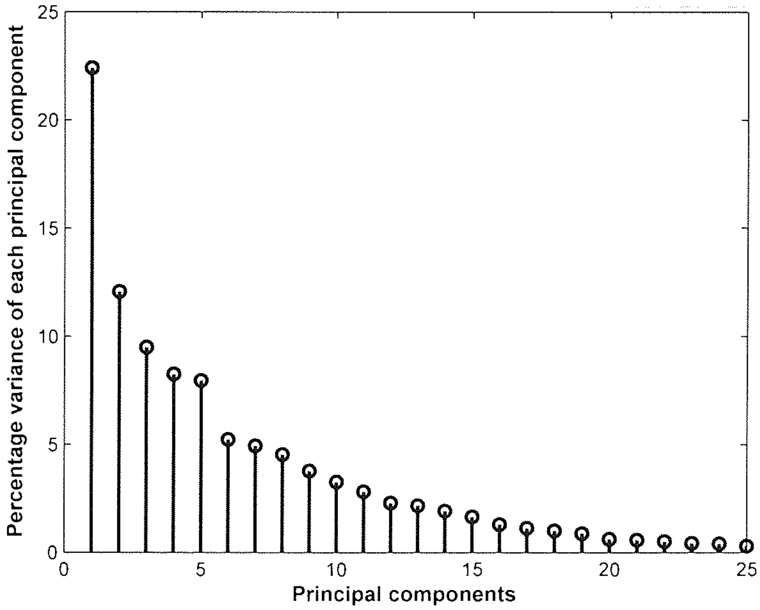


Figure 2.7: Percentage variance of each principal component of **X**.

where $i = 0, \ldots, p-1$.

$$\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \ldots; \mathbf{x}_p]'. \tag{2.8}$$

X with it's mean subtracted is input to the PCA. **Y** represents the eigenvectors of the covariance matrix $\mathbf{X}\mathbf{X}^T$, and C is the representation of **X** in the principal component space.

$$C = \mathbf{Y}\mathbf{X}. \tag{2.9}$$

The eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlations in the data set.

Figure 2.7 is the plot of the percentage variance of each principal component of **X**. The first five principal components contain nearly 60% of the total variance of the input data. Hence embedding the watermark in any of these five principal components preserves the watermark after a compression attack. The 3^{rd} principal component is chosen as it is robust against compression and it does not contain the maximum information of **c**. Any distortions to **c** along this direction leads to an acceptable value of $\mathcal{D}_{Emb}$. **k** is hence chosen from the representation of **c** over the 3^{rd} principal component. Figure 2.8 is the spectra of the third principal component. It contains a large power in the low frequency range supporting our argument that it is resilient to compression which typically removes high frequency components.

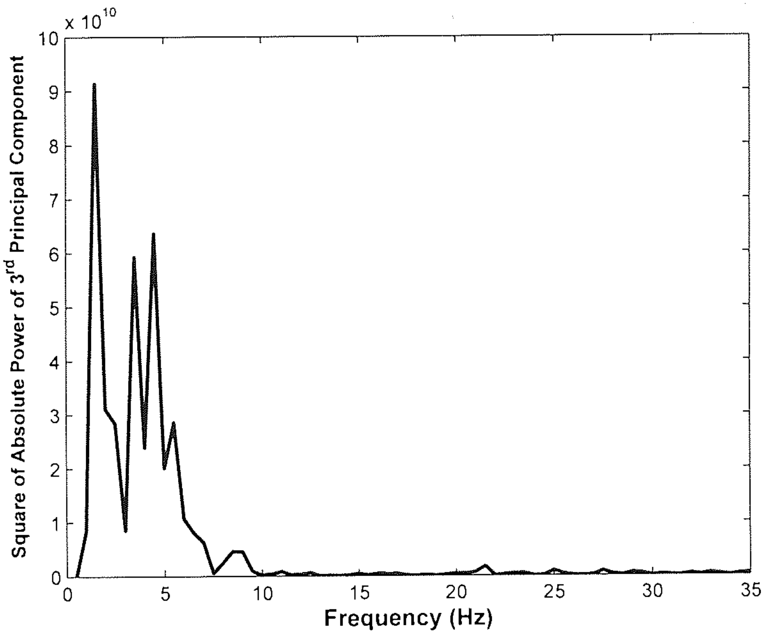


Figure 2.8: Frequency content of the 3^{rd} principal component.

The watermarked cover $\tilde{\mathbf{c}}$ can be obtained from the watermarked principal components as

$$\mathbf{Y}^{-1}\tilde{C} \rightarrow \tilde{\mathbf{c}}. \tag{2.10}$$

At the decoder let us assume the principal components of $\hat{\mathbf{X}}$ are derived by applying the PCA to $\hat{\mathbf{X}}$ derived from $\hat{\mathbf{c}}$.

$$\hat{\mathbf{X}} \xrightarrow{PCA} \acute{C}. \tag{2.11}$$

The principal components thus obtained are, $\acute{C}$ where $\acute{C} \not\approx \tilde{C}$. The embedded watermark is thus lost. In order to retrieve the embedded watermark the basis vectors $\mathbf{Y}$ derived at the encoder are required at the decoder.

2.1.5 Independent component analysis

ICA is a simple linear transformation [42]. Given p observation vectors

$$\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \ldots; \mathbf{x}_p]' \tag{2.12}$$

the ICA estimates l statistically independent source vectors.

$$\mathbf{S} = [\mathbf{s}_1; \mathbf{s}_2; \ldots; \mathbf{s}_l]' \tag{2.13}$$

and the mixing matrix $\mathbf{A}_{p\mathrm{x}l}$ where $l \leq p$, such that

$$\mathbf{X} = \mathbf{AS}. \tag{2.14}$$

As a watermarking principle, the message is embedded in one, or several of the independent sources treated as basis vectors spanning the space. Bounkong et al. in their paper [16] summarise the advantages of ICA compared to other transforms for watermarking applications.

- Statistical independence of the resulting sources.
- Estimates of one source provides no information of other sources.

They underline two reasons why using independent components is favourable for watermarking: the first is that the same framework may be applied to $D = 1,2,3$-dimensional covertexts (medical data), and so the same approach can be applied across different data modalities; The second is that in the context of watermarking, ICA allows the maximization of the information content and minimization of the induced distortion by decomposing the covertext into statistically independent sources. Information theoretical analysis presented in [67] also shows that the information hiding capacity of statistically independent sources is maximal.

Let $\check{\mathbf{S}}$ represent the l unknown independent sources and $\check{\mathbf{A}}$ the unknown random mixing matrix such that $\mathbf{X} = \check{\mathbf{A}}\check{\mathbf{S}}$. For the ICA to obtain a good estimate of $\check{\mathbf{S}}$ and $\check{\mathbf{A}}$, $\mathbf{S}$ and $\mathbf{A}$ respectively, the rows of $\mathbf{X}$ representing the number of observation signals $p \geq l$. With a one-dimensional observation $\mathbf{c}$ of length N_{samp} this condition is not satisfied. $\mathbf{X}$ can be constructed in one of two methods from $\mathbf{c}$ described below.

- Non-overlapping segments of $\mathbf{c}$ of length N_{nov}

  $$\mathbf{x}_i = \mathbf{c}[(i * N_{nov}) + 1, \ldots, (i-1) * N_{nov}]. \tag{2.15}$$

  where $i = 0, \ldots, p-1$.

- Delay Embedding of $\mathbf{c}$ with delay of one sample between two successive delay vectors, $\mathbf{d}_i$ and $\mathbf{d}_{i+1}$, embedding window size p,

  $$\mathbf{d}_1 = \mathbf{c}[1, 2, \ldots, p], \tag{2.16}$$

  $$\mathbf{d}_i = \mathbf{c}[i, \ldots, p+i-1], \tag{2.17}$$

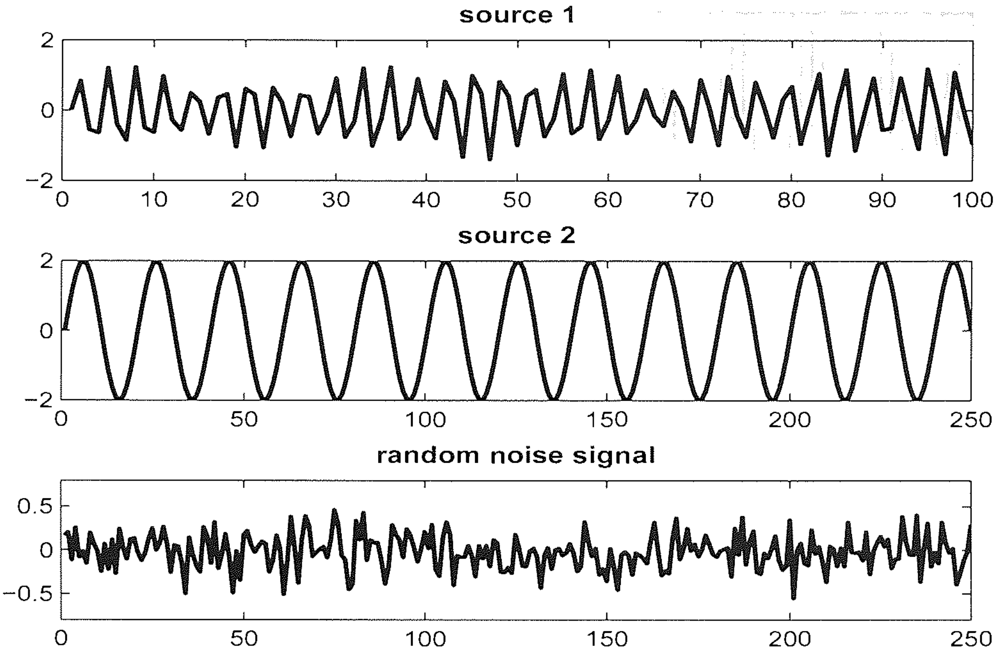


Figure 2.9: Synthetic data representing three independent sources.

where $i = 2, \ldots, N_{ov}$. The number of delay vectors $N_{ov} = \mathrm{N}_c - \mathbf{EmbWin} + 1$.

$$\mathbf{X} = [\mathbf{d}_1; \mathbf{d}_2; \ldots; \mathbf{d}_{N_{ov}}]'. \tag{2.18}$$

ICA applied to synthetic data

A demonstration with the help of synthetic data is provided to depict the result of applying ICA to an input matrix generated from a one-dimensional signal using equal non-overlapping segments and, equal overlapping segments obtained using the delay embedding method [101].

Figure 2.9 is an example set of three signals generated independently. Source 1 is a Gaussian pulse overlapped with a sine wave of 80Hz, source 2, a sine wave of 12.5Hz and source 3, a low power random noise signal. These signals are mixed using a 3x3 matrix of random values to obtain three mixed signals called observations. One of the observations is shown in figure 2.10. The frequency spectra of the observation signal and the three independent signals are shown in figure 2.11 (anticlockwise from lower right).

In order to obtain a comparison between the two methods of constructing the input matrix **X** the size of the embedding window **EmbWin** was calculated. The window size should be at least as large as the slowest frequency signal for the ICA to capture the

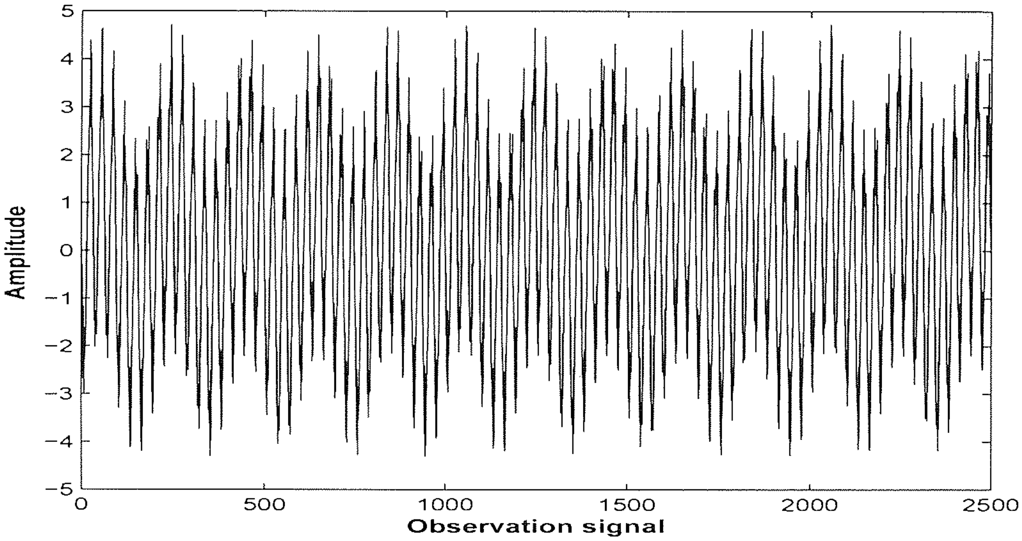


Figure 2.10: One of the mixed observation signals.

underlying dynamics. Let f be the smallest frequency signal of the one-dimensional observation, in this case 12.5Hz and f_s the sampling frequency of the observation, 250Hz.

$$\mathbf{EmbWin} = \frac{1}{f} * f_s. \tag{2.19}$$

The value of **EmbWin** thus obtained is 20. The observation signal of length 2500 samples when segmented using a window length of 20 gives 125 segments. Thus the input matrix to the ICA is $\mathbf{X}_{20\times 125}$. The estimated sources are as shown in figure 2.12. Comparing with the figure 2.9 it can be seen that the estimated sources do not represent the actual sources.

The sources shown in figure 2.13 are obtained by applying the ICA to an input matrix generated from **c** using the delay embedding method, with a delay of 1 sample and **EmbWin**=20. It can be observed that the estimated sources are a close approximation of the underlying sources though some of the sources are replicated.

The delay embedding method of estimating sources is hence a better technique of estimating the underlying sources.

ICA applied to single channel EEG data

The delay embedding method of creating the input matrix **X** from one-dimensional EEG, **c** is used. The ICA is applied to **X** to estimate the sources and the probable mixing matrix.

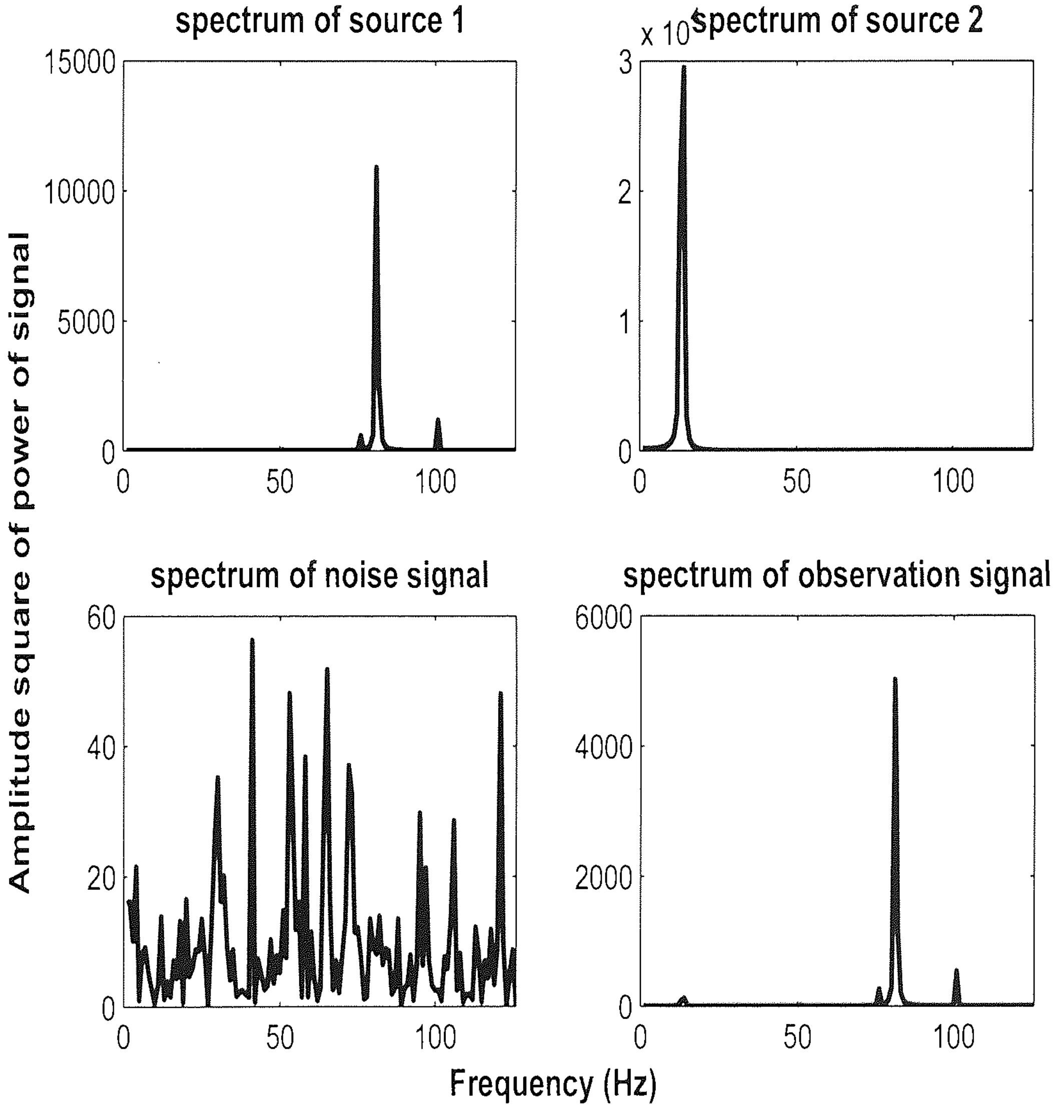


Figure 2.11: Frequency spectrum of the observation signal and the three sources respectively (anticlockwise from lower right).

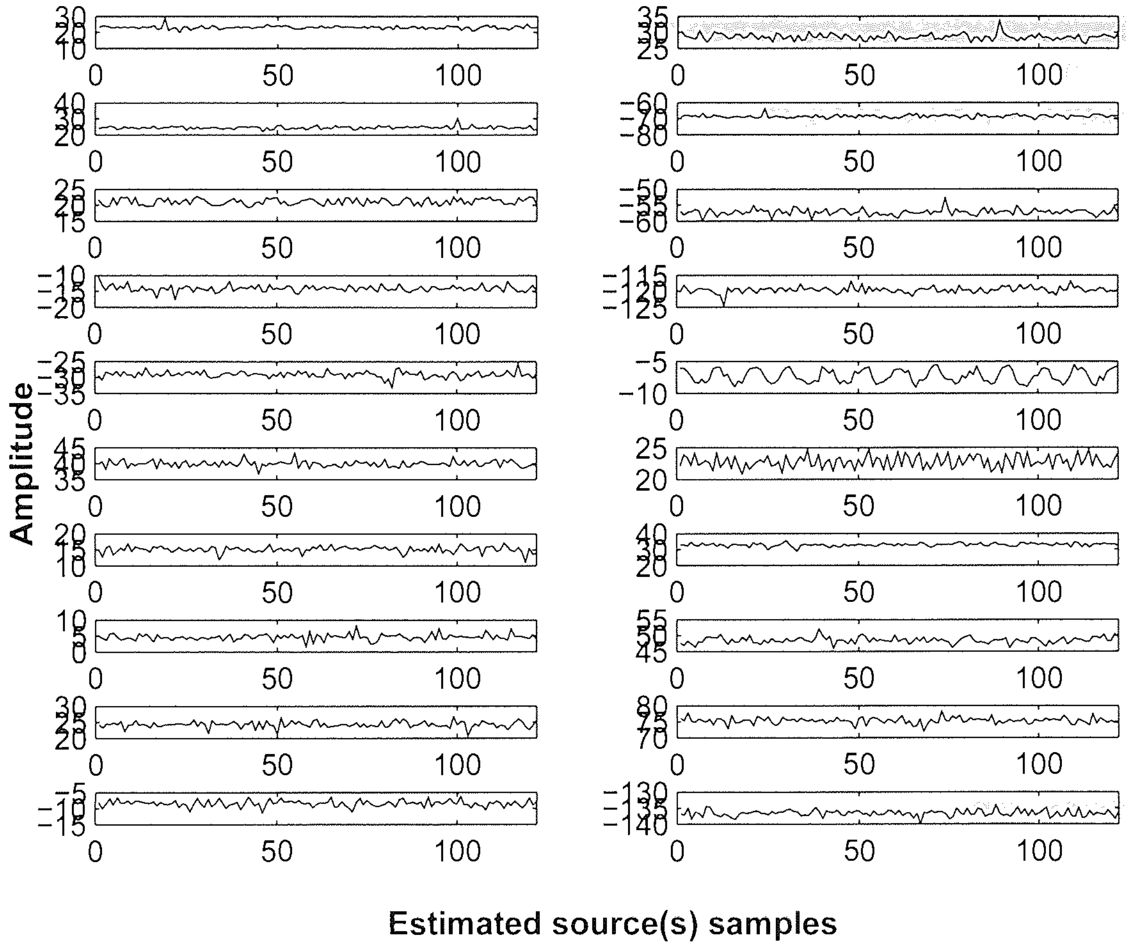


Figure 2.12: Sources estimated by the ICA when the input matrix is constructed from non-overlapping segments.

The separating matrix **W** is the inverse of the mixing matrix **A**.

$$[\mathbf{S},\mathbf{W}] \xleftarrow{ICA} \mathbf{X}. \tag{2.20}$$

$$\mathbf{W} = inv(\mathbf{A}). \tag{2.21}$$

The rows of **W** represent the independent components. **S** is the projection of **X** over this independent component space. One of the sources which represents the low and middle frequency components of **c** (see figure 2.14), $\mathbf{s}_{wm}$ is used and **k** is chosen randomly from samples of $\mathbf{s}_{wm}$ to embed the watermark.

$$\mathcal{F}_m(\mathbf{s}_{wm}(\mathbf{k}), WM) \rightarrow \tilde{\mathbf{s}}_{wm}. \tag{2.22}$$

Multiplying **A** with the matrix containing the watermarked source $\tilde{\mathbf{s}}_{wm}$ and the remaining

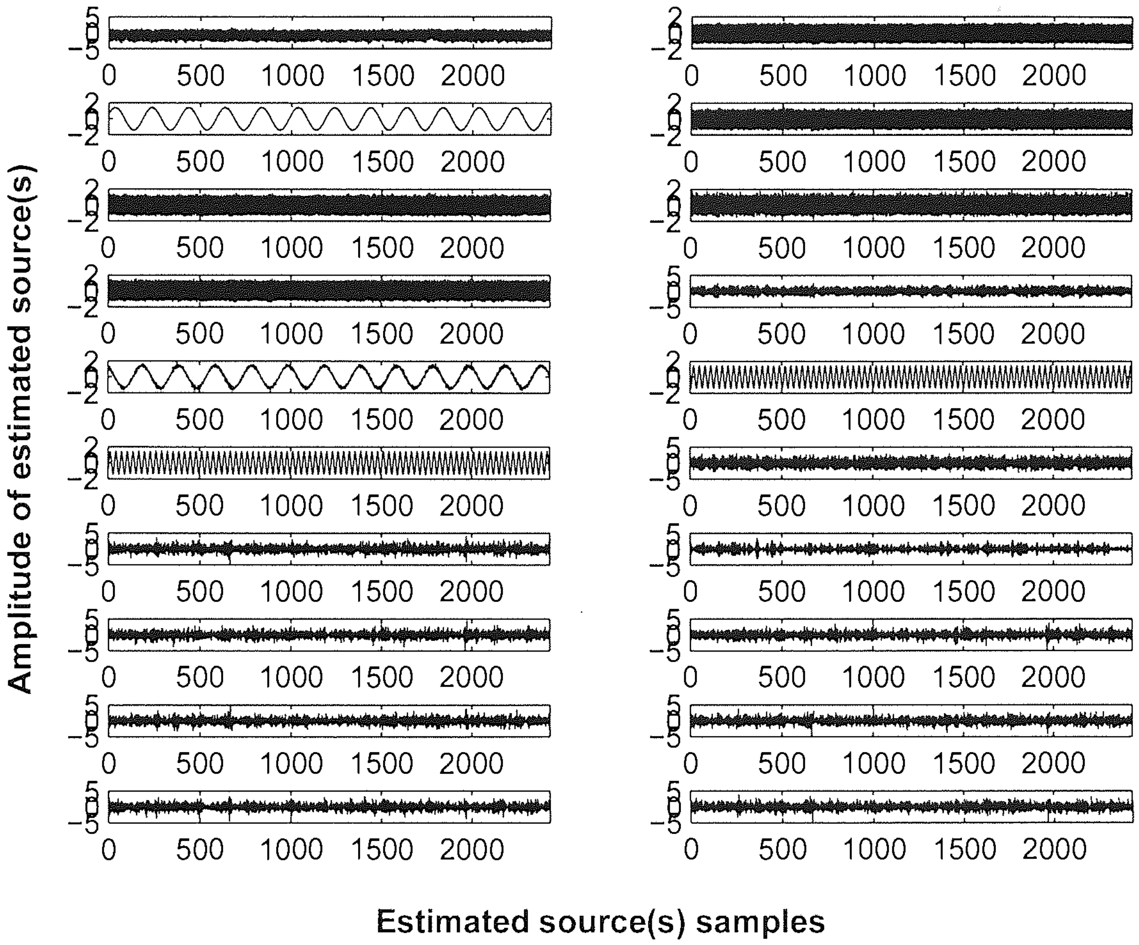


Figure 2.13: Sources estimated by the ICA when the input matrix is constructed from overlapping segments.

unwatermarked sources $\tilde{\mathbf{S}}$, $\tilde{\mathbf{X}}$ is obtained.

$$\tilde{\mathbf{X}} = \mathbf{A} * \tilde{\mathbf{S}}. \tag{2.23}$$

Reconstructing the one-dimensional $\tilde{\mathbf{c}}$ from $\tilde{\mathbf{X}}$ requires careful reordering of the samples. In the delay embedding method with a delay of one element, the elements of $\mathbf{X}$ on the diagonal connecting from top right to bottom left are numerically the same. This remains true for $\tilde{\mathbf{X}}$ except the values on columns represented by $\mathbf{k}$. Hence $\tilde{\mathbf{c}}$ can be reconstructed from $\tilde{\mathbf{X}}$ in one of two methods described below.

- Reconstruction of $\tilde{\mathbf{c}}$: Method 1

$$\tilde{\mathbf{c}}_i = \begin{cases} \mathbf{c}_i & \text{for i} \neq [wm + \mathbf{k}\text{-}1]\ , \\ \tilde{\mathbf{X}}_{wm_{\mathbf{k}}} & \text{for i} = [wm + \mathbf{k}\text{-}1]\ . \end{cases} \tag{2.24}$$

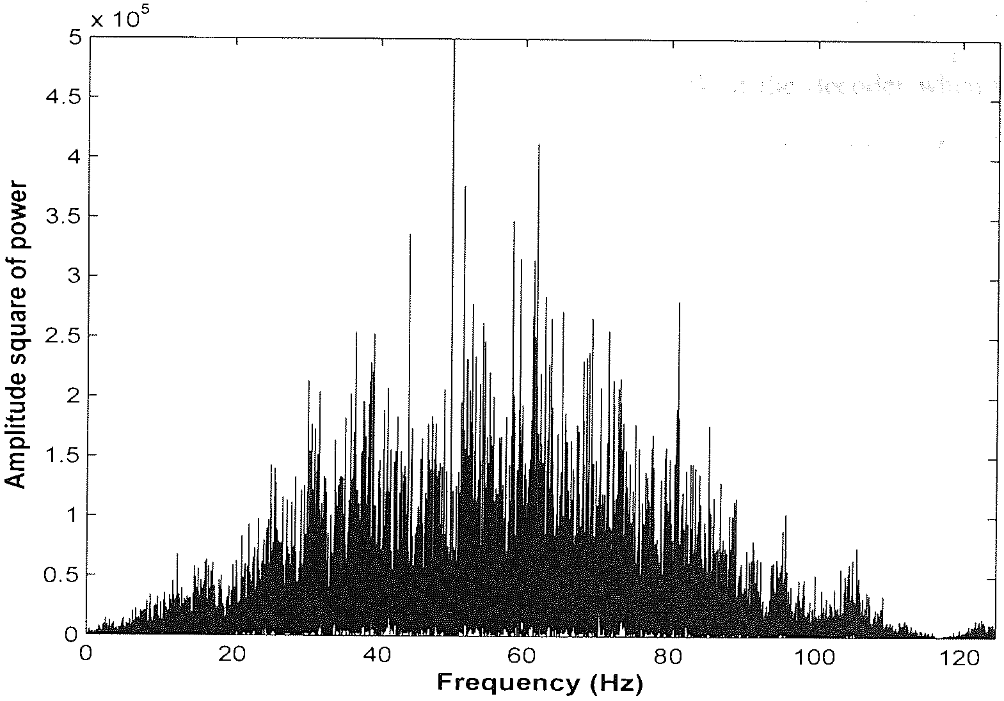


Figure 2.14: Frequency Content of One of the Estimated Sources.

This method demonstrates the advantages of the ICA mentioned above. Recovery of the embedded watermark is possible only when the ICA estimates statistically independent sources and the watermark is embedded in one such source. But the ICA does not estimate statistically independent sources from every one-dimensional time series data. This is because in most of the one-dimensional data (EEG) the actual number of the underlying independent sources is unknown. Secondly, the contribution of each independent source to **c** might not be equal. Lastly the amount of noise in the observation signal is not known. The ICA in most cases estimates interesting components and not necessarily statistically independent components.

This method of reconstruction provides a low value of $\mathcal{D}_{Emb}$ but does not always result in zero reconstruction error of the watermark at the decoder.

- Reconstruction of $\tilde{\mathbf{c}}$: Method 2

Since the sources are not statistically independent in most cases but merely interesting, modification of one source disturbs the other sources. Secondly all the elements across the diagonals connecting from top right to bottom left are the same so only one element across each such diagonal may be modified. When more than one element across the diagonal is watermarked parts of the embedded watermark will be lost at the embedder itself.

To maintain zero error reconstruction of the watermark at the decoder when the attack η=0, the following procedure is followed: **S** is divided into blocks of p columns each. Let $N_{b\mathbf{S}}$ represent the number of blocks of **S**. If p is a factor of N_{ov}, $N_{b\mathbf{S}} = N_{ov}/p$ else $N_{b\mathbf{S}} = (N_{ov})/p+1$.

$$\mathbf{S} = [\mathbf{S}(1:p,1:p)\mathbf{S}(1:p,p+1:2p)\ldots\mathbf{S}(1:p,N_{ov}-p+1:N_{ov})] \qquad (2.25)$$

Let $\mathbf{k_s}$ represent the elements of $\mathbf{s}_{wm}$ that could be watermarked. One element of $\mathbf{k_s}$ is chosen from one of the blocks of **S** such that $\mathbf{k_s}_i - \mathbf{k_s}_{i+1} = p$. **k** is chosen from $\mathbf{k_s}$ and is therefore limited by the number of blocks. $\tilde{\mathbf{X}}_{p\times N_{ov}}$ is reconstructed from $\tilde{\mathbf{S}}$ watermarked as mentioned above. $\tilde{\mathbf{c}}_i$ is obtained from $\tilde{\mathbf{X}}_{p\times N_{ov}}$ as follows.

$$\mathbf{X} = [\mathbf{X}(1:p,1:p)\mathbf{X}(1:p,p+1:2p)\ldots\mathbf{X}(1:p,N_{ov}-p+1:N_{ov})] \qquad (2.26)$$

$$\tilde{\mathbf{c}}(n+(0:p-1)) = \tilde{\mathbf{X}}(1:p,n) \qquad (2.27)$$

where $n = 1,2,\ldots,N_{b\mathbf{S}}$.

This method of reconstruction provides a low $\mathcal{R}$ as all the samples of $\mathbf{s}_{wm}$ cannot be watermarked, and the value of $\mathcal{D}_{Emb}$ is higher than in the reconstruction method 1 but the watermark can be retrieved at the decoder with zero error for limited η. This method satisfies the criteria of a biomedical watermarking system to a compared to method 1 in terms of watermark retrieval with no error. Hence all the experiments are conducted using this method.

As in the case of PCA, **W** is data dependent. Hence the same **W** is required at the decoder to retrieve the embedded watermark [35].

2.2 Watermark Embedding Techniques

Though many different empirical and standard methods of watermark embedding exist one watermark embedding technique, utilising lattice structures, Quantisation Index Modulation (QIM) based embedding technique is one of the most popular [2]. The watermark embedding process is depicted in figure 2.15.

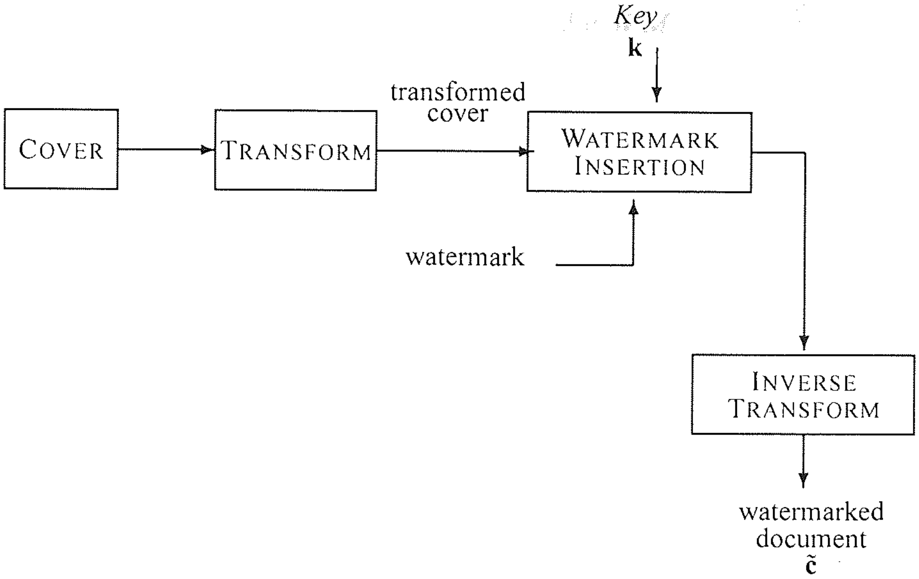


Figure 2.15: Block diagram of a generic watermark embedding system.

2.2.1 Quantisation method

'QIM' refers to embedding information by first modulating an index or sequence of indices with the embedded information and then quantising the host signal with the associated quantiser or sequence of quantisers [20]. Quantisers are defined as a class of discontinuous, approximate-identity functions. δ defines a quantisation index representing a scalar quantiser.

For a binary watermark WM, QIM generates the watermarked host data $\tilde{\mathbf{c}}$ as

$$\tilde{\mathbf{c}}(k_i) = \begin{cases} \delta N_e, & \text{if } WM_i = 0 \\ \delta N_o, & \text{if } WM_i = 1 \end{cases} \tag{2.28}$$

where N_e and N_o are respectively even and odd integers and $\mathbf{c}(k_i)$ is quantised to the nearest δN_e or δN_o. In QIM based watermark embedding the parameter δ defines the position of the watermark on the trade-off triangle. For the embedded watermark to be robust against an attack the value of δ should be large and to comply with the distortion constraint for $\mathcal{D}_{Emb}$, the function $\mathcal{F}_m$ should be close to identity. At the decoder the nearest level decoding method is adopted to estimate the probable WM.

$$\widehat{WM}_i = \begin{cases} 0, & \text{if } \frac{c_j}{\delta} \sim 0, \\ 1, & \text{if } \frac{c_j}{\delta} \sim 1. \end{cases} \tag{2.29}$$

2.3 Preliminary Experiments and Results

The EEG segments were drawn from 3 different datasets. One EEG showed no abnormalities. One EEG represented photosensitive epilepsy and the other was recorded for a patient undergoing epileptic seizures (cause was not specified). A total of 70 different one-dimensional EEG signals were used to conduct each of the experiments. Since each EEG signal contained different band limited signals, to maintain uniformity the number of quantisation levels used for QIM embedding and the number of watermark bits embedded were kept constant. The bit error rate, *dist* and the distortion varied for each EEG. The difference in the bit error rates were less than 5% and the distortion around 1% of the dynamic range of the EEG signal.

Each EEG segment considered is of 100s duration, sampled at 250Hz with each sample represented as a 16bit unsigned integer. The various transforms of DFT, DCT, DWT, PCA and ICA are applied to the single channel EEG. For the transforms of DFT, DCT and DWT, **c** is the one-dimensional EEG of 100s. For PCA and ICA the one-dimensional EEG is transformed into a matrix. As explained in [96] the EEG is segmented into non-overlapping segments to form an input matrix for the PCA. In the case of ICA, since non-overlapping segments constituting the input matrix do not result in sources which are approximately close to the actual sources, the dynamical embedding method is applied. The transformed co-efficients are watermarked as explained in the respective sections on each transform for different rates of $\mathcal{R}$. The value of δ used is maintained constant. The values of C obtained for each transform have different dynamic ranges. Therefore for experimental purposes and to obtain a comparison of the different methods δ is chosen based on a fixed number of quantisation levels equal to 195. The watermark detection is as shown in figure 2.16.

Medical signals are not likely to be subject to the same type of malicious attack as, say, downloaded music or video files. However attacks such as pre-signal processing, or downsampling of large data files to allow more efficient data transmission could be an issue. The robustness of the watermark is verified against different attacks such as low

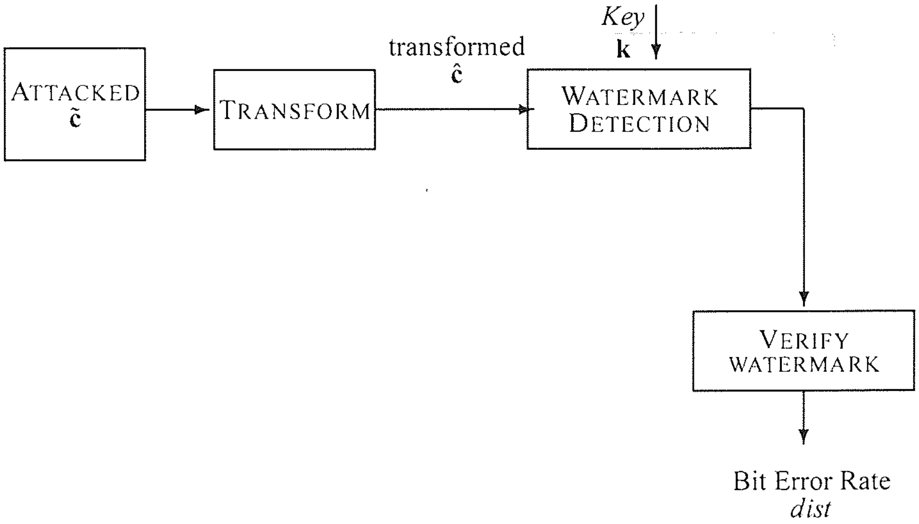


Figure 2.16: Block diagram of a generic watermark detection system.

pass filtering, addition of Gaussian noise, different sampling rates and compression. The decoding for each watermarking technique is conducted as explained in the section on QIM and the resulting value of the bit error rate, *dist* is plotted for each method.

Figure 2.17 is a plot of *dist* versus $\mathcal{R}$ of the embedded WM. This value of $\mathcal{R}$ is a percentage of the total length of C available to embed the WM.

The message can be reconstructed without error in the case of DFT and DCT since these transforms provide the spectral information of **c** and the samples of **k** chosen for these methods are clearly within the passband of the filter used as an attack. C in the PCA method has a wide-band frequency representation but has more low frequency components compared to the value of C obtained in the DWT and ICA methods. Though the DWT and the ICA method do not provide the spectral information of **c** leading to a higher *dist* compared to the DFT and DCT methods they are advantageous in an eHealth system application. The DWT and ICA methods produce mutliple non-interfering signals required to embed multiple watermarks. This feature cannot be obtained by applying the DFT or the DCT transform to a one-dimensional signal.

Figures 2.18, 2.19, 2.20, 2.22 show the bit error rates of the retrieved watermarks (5 watermarks are embedded using DFT, DCT, DWT, PCA and ICA based methods) for different sampling rates, low pass filtering, addition of Gaussian noise and compression

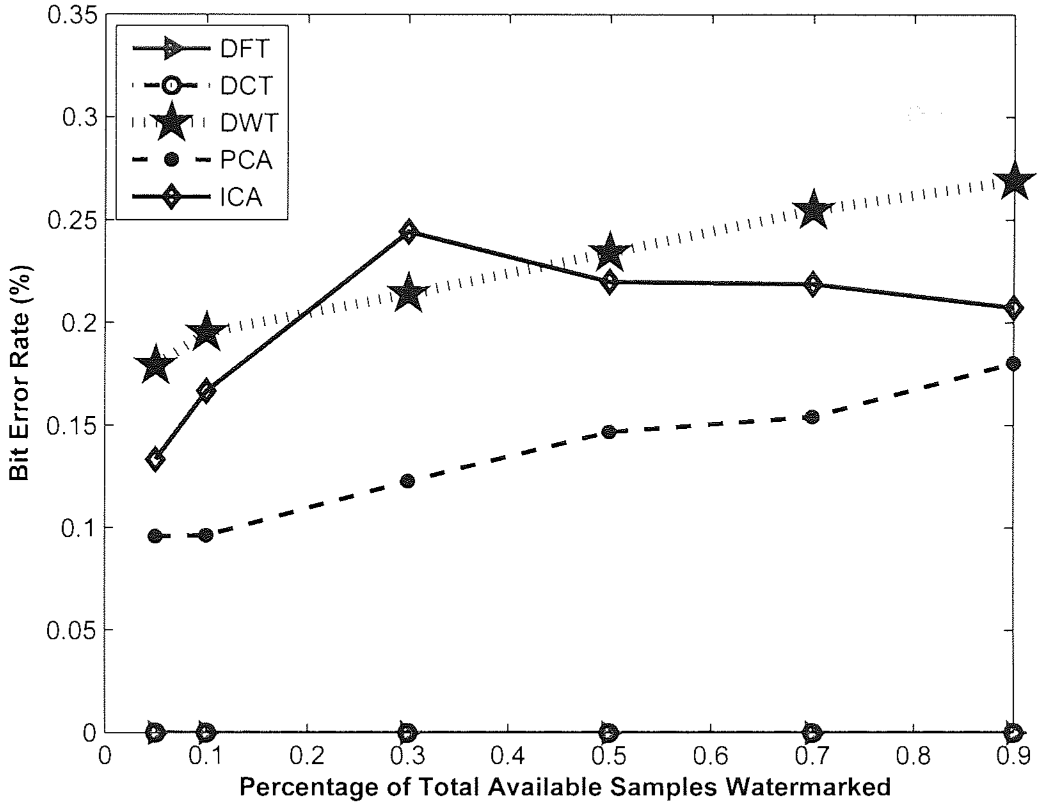


Figure 2.17: Comparison of robustness of *WM* embedding methods (DFT, DCT, DWT, PCA and ICA) for different $\mathcal{R}_c$ against a filter attack.

attacks respectively. The x-axis represents the ratio of the watermark signal to the strength of the attack. It is given by $\frac{\mathbf{c}-\tilde{\mathbf{c}}}{\tilde{\mathbf{c}}-\hat{\mathbf{c}}}$. It is represented as watermark to noise ratio. The y-axis is the bit error rate defined as *dist*. It can be seen that the bit error rate is less than 15% for the sampling, compression and noise attacks. This error rate could be corrected with the use of error correcting codes [53]. Applying error correction and verifying the best error correcting code is beyond the work produced in this thesis and is therefore not investigated. The performance of the PCA and ICA methods is greater than 15% for the low pass filter attack.

The EEG signals used for the experiments were encoded using 16 bit unsigned integer format. The watermarked EEG was requantised such the number of bits per sample varied from 6 to 14. The result of this decreasing encoding rate on the embedded watermark was verified. The results are shown in figure 2.18. The bit error rate was plotted against the watermark to noise ratio (noise here represents the distortion to the watermarked EEG being encoded using a lower number of bits). It can be seen that the performance of all the transform based methods is nearly the same.

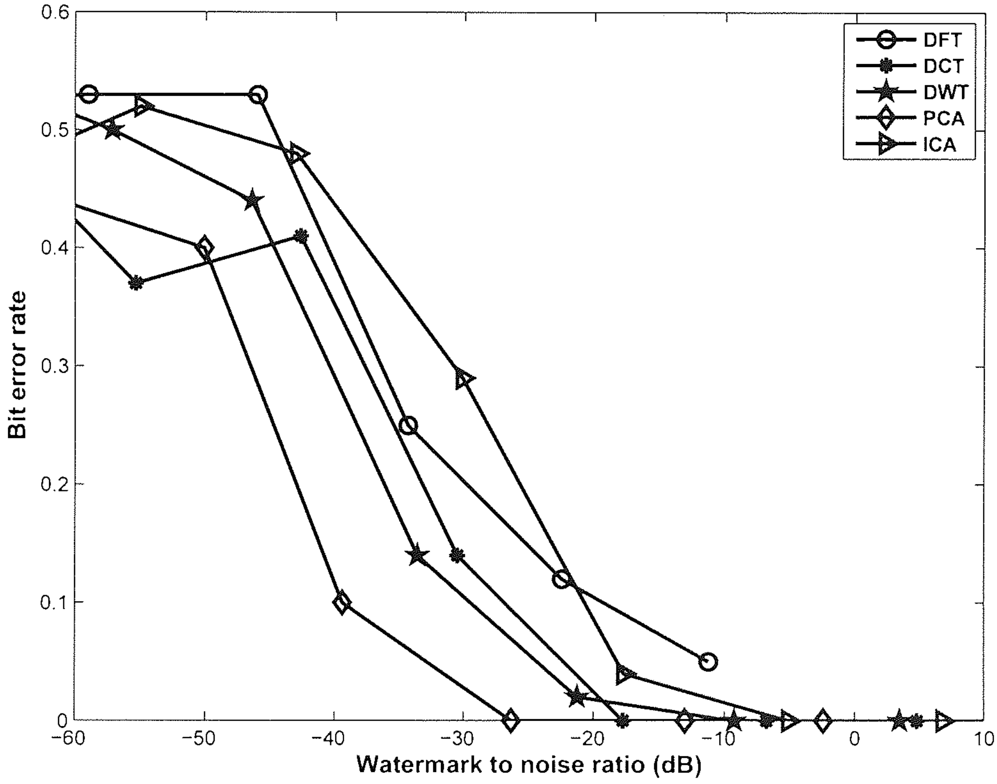


Figure 2.18: Comparison of robustness of *WM* embedding methods (DFT, DCT, DWT, PCA and ICA) for different sampling rates.

The low pass filter used is a Butterworth filter of order 4. The Butterworth filter was used as it has a good allround performance and better rate of attenuation [82]. The order of the filter was varied from 2 to 10 and the results noted. There was a difference of less than 5% in the bit error rate obtained for different order filters. Since the samples used to embed the watermark are chosen randomly, different iterations give slightly varying results but as the difference is less than 5%, it has been ignored. The results (figure 2.19) presented in this chapter are for a Butterworth filter of order 7. The performances of the DFT, DCT and DWT based methods are similar. This is because the transform co-efficients of these three methods contained similar spectra. The performance of the the ICA and PCA based methods is worser because the spectra of the co-efficients used to embed the watermark in these methods contained higher frequencies.

Additive Gaussian noise is also added to the watermarked EEG to test the robustness of the embedded watermark. The variance of the Gaussian noise is increased from 0 to

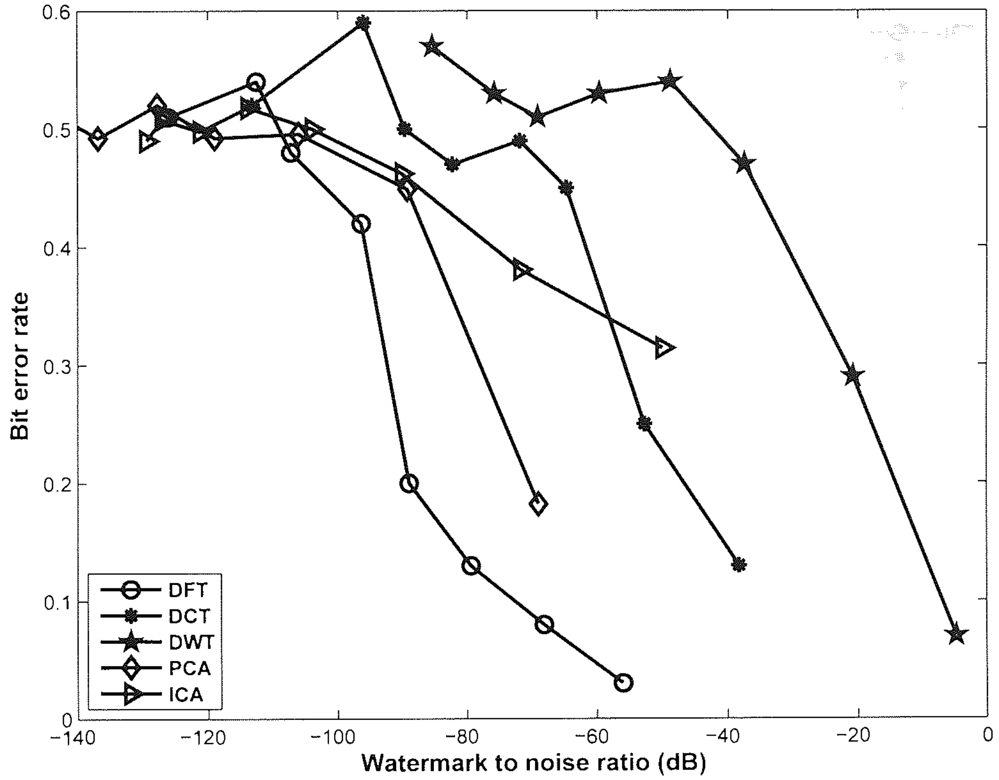


Figure 2.19: Comparison of robustness of WM embedding methods (DFT, DCT, DWT, PCA and ICA)against low pass filtering attack for different cut-off frequencies of a low pass Butterworth filter.

3.5. From figure 2.20 it can be seen that the performance of the DFT based method is comparatively worser. This is due to the spectra of the DFT co-efficients used to embed the watermark having lower frequencies compared to the co-efficients obtained for the DCT, DWT, PCA and ICA based methods.

EEG compression techniques have been described in [5, 14, 63]. Hence in order to study the effects of compression, the single channel EEG embedded with the watermarks is modelled as an AR process of different orders from 1 to 6. Let N represent the order of the AR process. The Levinson-Durbin method of recursion, *LPC* [7] is used to compute an N-th order forward linear prediction polynomial represented by A. 'The Levinson-Durbin method provides a big saving in the number of operations (multiplications or divisions) and storage locations compared to other standard methods such as the Gauss elimination method' [39]. Hence we use the Levinsion-Durbin method of representing the EEG as an AR process.

$$\mathrm{A} \overset{LPC}{\longleftarrow} \tilde{\mathrm{c}}. \tag{2.30}$$

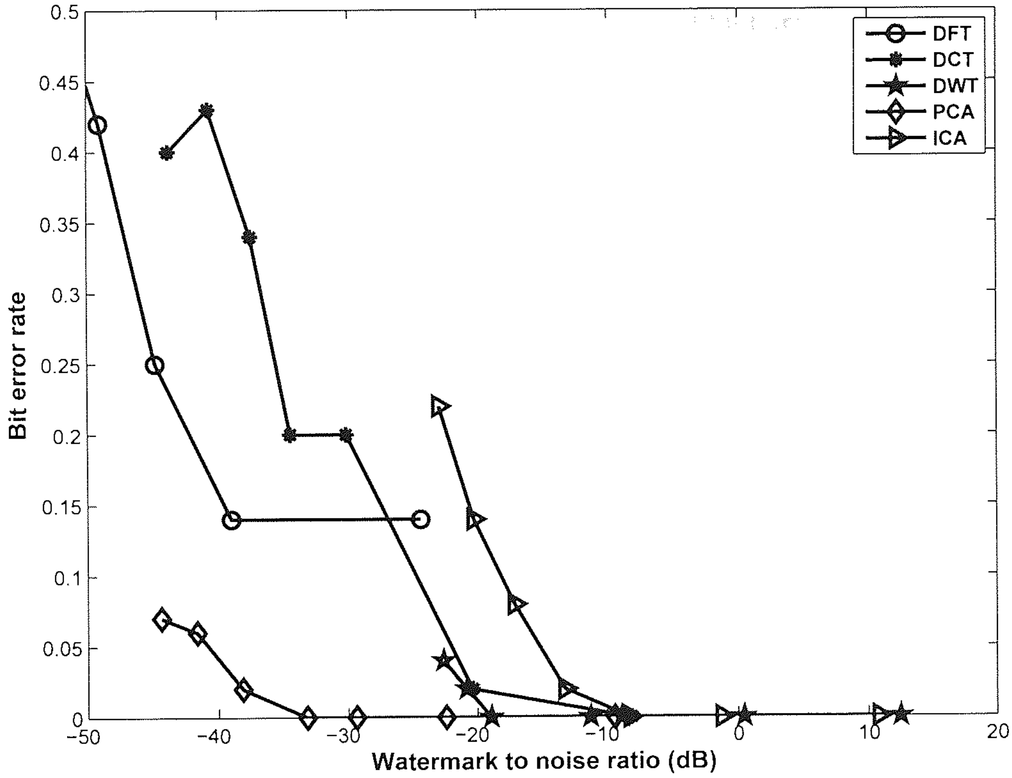


Figure 2.20: Comparison of robustness of *WM* embedding methods (DFT, DCT, DWT, PCA and ICA) for different levels of additive noise.

Let $\tilde{\mathbf{c}}_p$ represent the n-th predicted future value of $\tilde{\mathbf{c}}(1,\ldots,n-1)$. Given polynomial A and N samples of $\tilde{\mathbf{c}}(1,\ldots,n-1)$, $\tilde{\mathbf{c}}_p(n)$ is obtained as follows.

$$\tilde{\mathbf{c}}_p(n) = -\mathrm{A}(2) * \tilde{\mathbf{c}}(n-1) - \mathrm{A}(3) * \tilde{\mathbf{c}}(n-2) - \ldots - \mathrm{A}(N+1) * \tilde{\mathbf{c}}(n-N). \qquad (2.31)$$

Let e_p represent the predicted error

$$e_p(n) = \tilde{\mathbf{c}}(n) - \tilde{\mathbf{c}}_p(n). \qquad (2.32)$$

The high valued 16 bits per sample $\tilde{\mathbf{c}}$ was converted to a representation of the low valued, variable length format e_p obtained from the AR process to obtain a saving in length.

Figure 2.21 shows the bit error rate obtained when the watermarked EEG is compressed by modelling it as an AR process of orders 1 to 10. The EEG signal was watermarked using the DWT based method (other transform based domain methods were also tested). The order of the AR process did not change the bit error rate significantly nor was there a particular order to the bit error rates. This is due the random selection of the samples used to embed the watermark. So the attack on the watermarked content was designed by representing the error co-efficients using different bit rates.

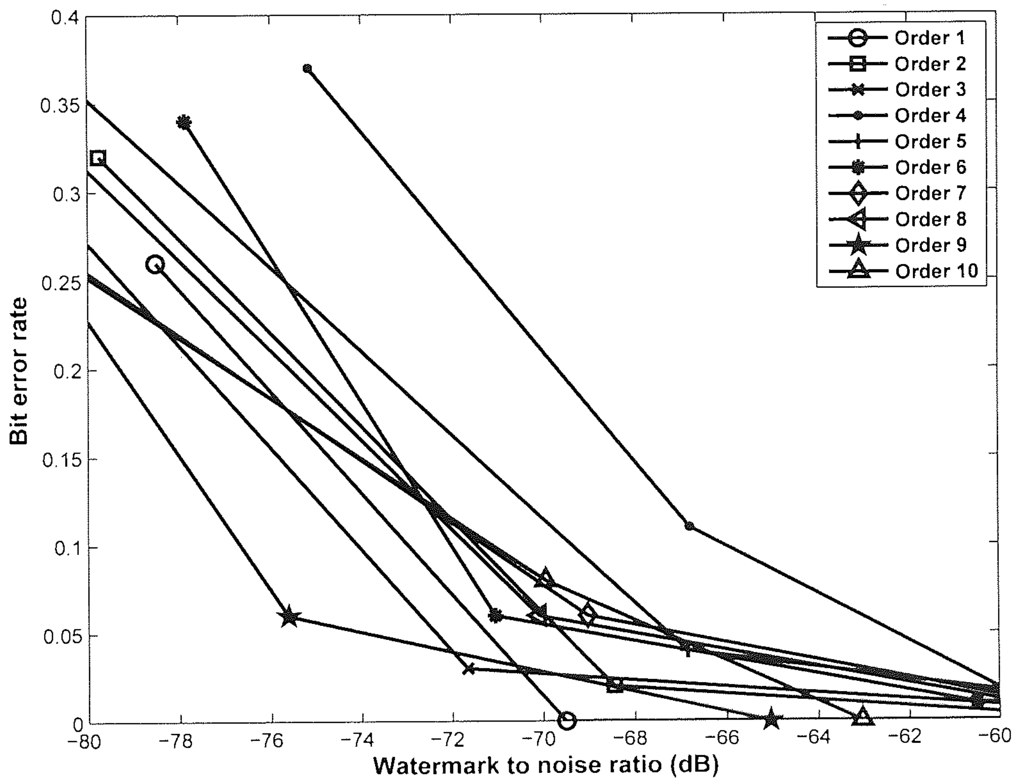


Figure 2.21: Comparison of robustness of WM embedding methods (DWT) for different levels of compression using LPC of different orders.

The error co-efficients were represented using different rates from 6 bits to 14 bits giving a compression of 37% to 87%. Since the compression obtained is based on the number of bits used to encode each sample of the error co-efficients, the order of the AR process used was 6. The bit error rate is obtained for different compression rates and is represented as the watermark to noise (noise here is the compression) ratio.

Figure 2.23 is the distortion to $\mathbf{c}$ due to the embedded WM. This distortion is plotted for the value of $\mathcal{R}$ of the WM recovered with smallest error rate. The distortion is calculated using the mean square error (discussed in chapter 1) between $\mathbf{c}$ and $\tilde{\mathbf{c}}$. The fidelity remains high for message embedding data rates of a few bits per second, and this translates to an insignificant perceptual change of the observed transmitted signal when reviewed by the clinician. The maximum distortion for any given transform equates to a few microvolts. Typical magnitudes of recorded EEG are in the range of tens–hundreds

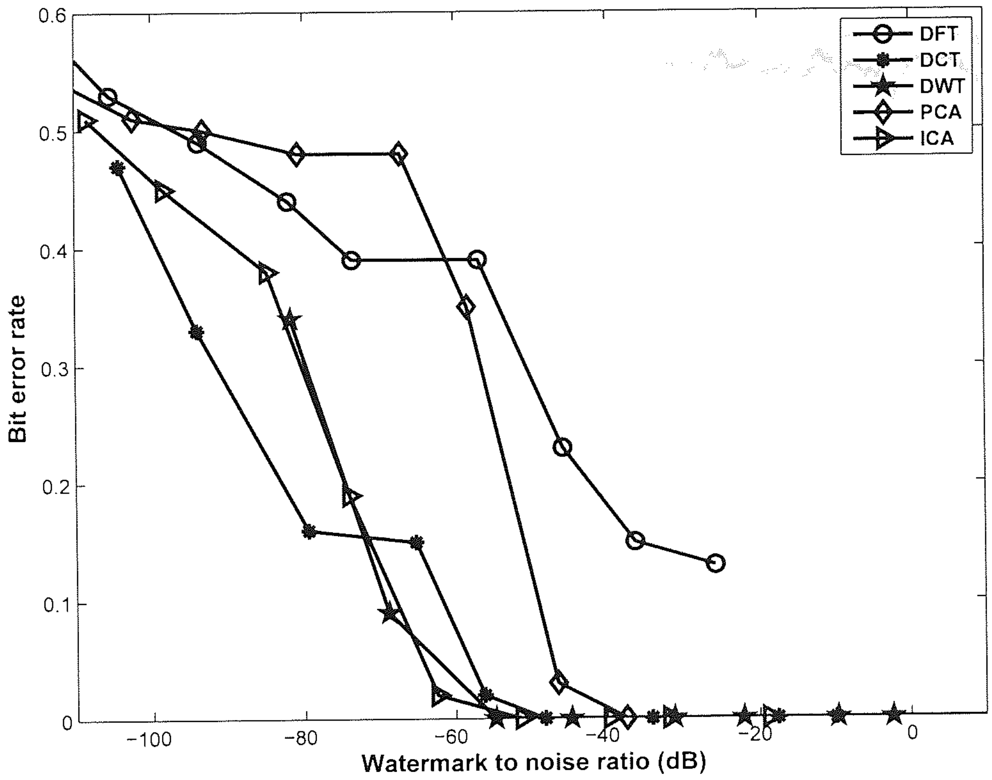


Figure 2.22: Comparison of robustness of *WM* embedding methods (DFT, DCT, DWT, PCA and ICA) for different levels of compression.

of microvolts.

Figure 2.24 depicts a histogram showing the distribution of the *mean absolute deviation* induced by running the watermarking process (ICA method) one thousand times. Again, this figure confirms that the distortion over a wide range of embeddings is restricted to a few microvolts.

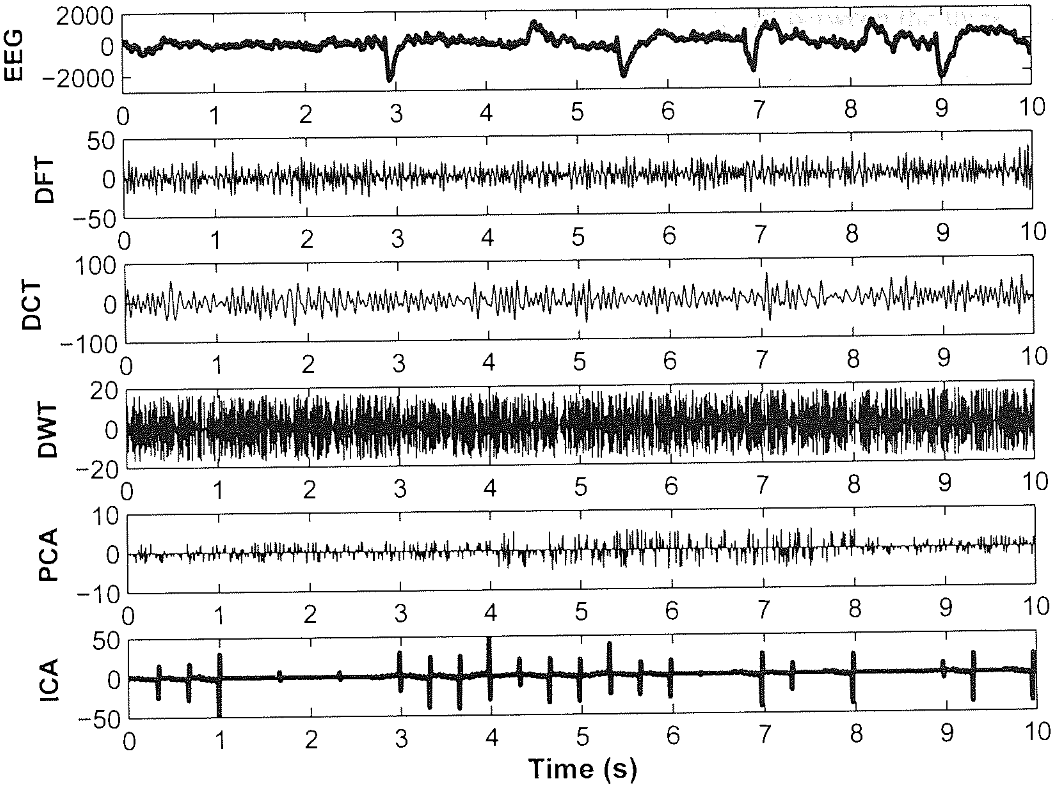


Figure 2.23: Distortion due to the embedded WM, comparison of different embedding methods.

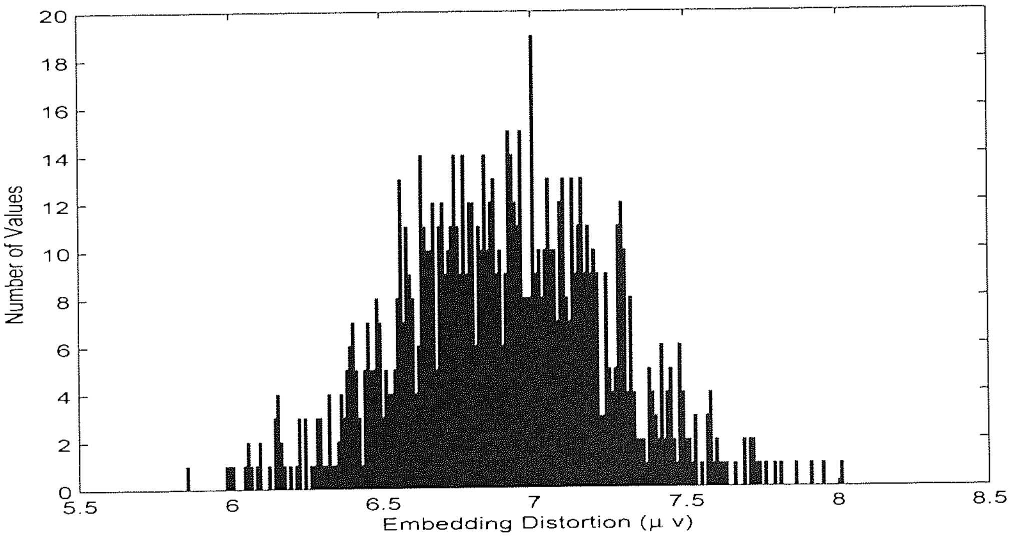


Figure 2.24: Histogram of the distribution of maximum distortions induced using a sample rate of $5bs^{-1}$ over many random realisations of the message embedding process. Distortion in the original data due to the embedding message at $5bs^{-1}$.

2.4 Conclusion

From figures 2.17, 2.18, 2.19, 2.20, 2.22, and 2.23, the trade-off between the three characteristics of robustness, imperceptibility and data rate are obvious. As the value of $\mathcal{R}$ is increased the robustness decreases. If the robustness of a method is high, the imperceptibility is low. From the discussion on the derivation of $\mathcal{C}$ using the various transforms it is realised that the possibility of embedding multiple watermarks of differing requirements of robustness and security in a single channel time series data is possible by using the DWT, PCA and ICA methods. The DFT and DCT methods provide a one-dimensional $\mathcal{C}$ for a one-dimensional $\mathbf{c}$. Embedding multiple watermarks in these domains needs careful segmentation of $\mathcal{C}$ so as to ensure that the multiple watermarks are not overwritten by one another. This problem does not exist in the embedding methods based on DWT, PCA and ICA as they provide multiple orthogonal/independent channels as output.

The transform co-efficients of DWT are fixed and independent of the data while both PCA and ICA are data dependent. The advantage of ICA over PCA is, the output of the ICA, the estimated sources are independent of each other and hence non-interfering. Embedding information in one of the source will not modify or distort the information content of the other sources.

In the next chapter the security of QIM based watermark embedding techniques (scalar and dither modulation(DM)-QIM) is investigated for watermarks embedded in the DWT and ICA based watermarking systems.

3 SECURITY OF SCALAR QIM AND DM-QIM BASED EMBEDDING TECHNIQUES

CONTENTS

In this chapter the established notion in watermarking literature, that the commonly used QIM based watermark embedding method is secure, will be challenged. In chapter 2, different signal processing techniques were applied on the watermarked cover, $\tilde{\mathbf{c}}$ to verify the robustness of the message. The robustness of the watermark is measured by applying an attack, η, which represents a signal processing technique applied to $\tilde{\mathbf{c}}$. The characteristics of the embedding function, $\mathcal{F}_{\mathrm{m}}$ is measured on the trade-off between the three properties of robustness, imperceptibility and data rate but the security offered by $\mathcal{F}_{\mathrm{m}}$ to the embedded message, $\mathbf{m}$ has rarely been explored in many works.

It was seen in chapter 2 that unlike the DFT, DCT and DWT where the transform coefficients are fixed, the PCA and ICA methods derive their basis vectors from the input data. In an information hiding application the hidden message, $\mathbf{m}$ distorts the cover work, $\mathbf{c}$. Therefore, the distortion to the cover work $\mathcal{D}_{Emb}$ is maintained to be under a threshold value. This threshold is defined based on the application of the watermark. The basis vectors derived in the PCA and ICA methods from $\mathbf{c}$ and the attacked watermarked cover, $\hat{\mathbf{c}}$ differ to an extent that the watermark retrieved at the decoder $\hat{\mathbf{m}} \not\sim \mathbf{m}$. Hence the basis vectors used at the embedder are required at the decoder to retrieve an estimate of the embedded message such that $\hat{\mathbf{m}} \sim \mathbf{m}$. This extra information which needs to be transmitted securely to the decoder is an extra payload but this extra payload could prevent illegal data snooping.

QIM is currently the most popular method of embedding information. There are nearly 770 papers on watermark embedding utilising the basic/modified versions of QIM techniques presented in the last 8 years [2]. The watermark embedding technique in chapter 2 was based on QIM techniques. In this chapter the security of QIM based embedding techniques to messages embedded in time series data is tested. We ask 'How close can an attacker get to recovering the message, given she knows the method and key factors such as segmentation blocking?' An attacker who intercepts $\hat{\mathbf{c}}$ can use their knowledge of similar $\mathbf{c}$ to derive their own estimates of the basis vectors and obtain the projections of $\hat{\mathbf{c}}$ over these basis vectors. Assuming that the attacker knows $\mathcal{F}_{\mathrm{m}}$, she constructs $\hat{\mathbf{m}}_{Att}$, but using her estimates of the projections. Will $\hat{\mathbf{m}}_{Att} \sim \mathbf{m}$? How sensitive is the ICA method to changes in $\mathbf{c}$? Will this sensitivity allow the discovery of the message by a third party having access to only $\hat{\mathbf{c}}$?

The main aim of this chapter is to investigate a secure method of providing patient privacy in the EPHR. Watermarking of biomedical data with the personal information in a medical record is currently the best possible method to ensure that the sensitive personal information is secure. Therefore assessing the security of the embedding method and finding the answers to the above questions is critical to any system implementation.

The embedding of the watermark is achieved in the transform domain of the time series data namely using DWT and ICA. Although a number of variants of the QIM method exist, the security of the two most commonly used of them, scalar QIM and Dither Modulation (DM) QIM are investigated. An investigation of the additional security of the embedded message due to the sensitivity of the recovered messages to slight variations in the structure of the independent components, or knowledge of which components have been modified - a topic not previously considered, and which augments any cryptographic approach for security will also be conducted.

3.1 Key Security

In an EPHR wherein the embedded messages (for example, details of personnel who have accessed the record) are used as a log of access, an unauthorised person who has accessed the record will attempt to delete her access details. Therefore finding the secret key used in the embedding of the messages makes it possible to erase the hidden message (watermark) without distorting the watermarked content to a large extent. Since the personal information relating an individual to an EPHR requires a high level of security, it can be further secured by encrypting it using suitable encryption methods such as RSA [89]. In the event, the attacker has gained access to the secret key used in the watermark embedding method and has retrieved the embedded message without error, if the watermarks were encrypted prior to the embedding then the unauthorised user (attacker) will need extra keys to obtain the true hidden message. There exists considerable literature on cryptography and exploring suitable encryption methods is beyond the scope of this thesis, hence we do not consider any encryption of the data.

The work presented in this chapter is based on the assumption that the cover data contains embedded messages. The security of hidden data is provided by the use of secret embedding keys. Kalker's [50] definition of watermarking security which he explains as the inability of unauthorised users to remove, detect and estimate, write or modify the raw watermarking bits, forms the basis on which the security of QIM based watermark embedding techniques are assessed in the work presented in this chapter. The concept of key security for QIM based watermarking has been investigated by various authors [50, 40, 10, 18, 76, 31, 75].

Overwriting the information in the watermarked cover to partially or completely destroy the original information is possible if $\mathcal{F}_{\mathbf{m}}$ is known. Even if the exact $\mathbf{k}$ is not known, it is possible to destroy $\mathbf{m}$ by randomly overwriting $\hat{\mathbf{c}}$ to a large extent. Cox et al [25] claim that $O(\sqrt{\mathbf{k}/ln\mathbf{k}})$ similar watermarks must be added to $\tilde{\mathbf{c}}$ to destroy the original watermark. In the work presented by them $\mathbf{k}$ represents the number of most perceptually significant frequency components of an image's discrete cosine transform used to embed the original watermark. The watermark used in their experiments is a sequence of real numbers drawn from a Gaussian distribution. But this method has a serious disadvantage. The possibility of the attacker's message destroying the usability of $\hat{\mathbf{c}}$ increases. Any benefit that the attacker may wish to gain will be lost. Hence a method of estimating $\mathbf{k}$ is necessary in order to destroy or overwrite $\mathbf{m}$ and still maintain the viability of $\hat{\mathbf{c}}$ for use. The security of the message content is considered to be assured since it lies hidden in the host signal distributed randomly and the random distribution pattern is known only to the owner of the host signal.

Cayre et. al. [18] state that it is possible to guess certain information about the secret key from the watermarked content. They define this term as information leakage and show that information leakage can be quantified by measures such as mutual information. Cayre et al [18] grouped the attacks by using the Diffie and Hellman methodology for security of cryptographic systems, on watermarked content as (1) Watermarked only attack (WOA) - wherein the attacker has access to a set of watermarked host data. (2) Known-message attack (KMA)- where the attacker has access to a set of watermarked content and the associate messages. (3) Known-original attack (KOA), where the attacker has access to both the watermarked content and the original unwatermarked content. Using tools from information theory various measures to estimate information leakage about the secret key from the observable data have also been discussed in the paper.

Cayre and Bas [31] state that 'an embedding function is key secure if it is impossible to estimate the secret key, even if the secret subspace to which the secret key belongs can be estimated. Knowledge of the secret subspace reduces the uncertainty of the secret key and therefore the security of the secret key is dependent on the number of possible keys that can be obtained in the subspace'.

An in-depth estimation of the secret key for spread spectrum based watermarking methods for the WOA class of attacks can be found in ([31]). Pérez-Friere et al [75] have presented an extensive analysis of lattice-based data hiding methods. The work presented in ([76]) is based on the assumption that the attacker has access to several copies of the data watermarked with the same secret key. For the DM-QIM watermarking technique, Bas and Hurri ([10]) showed how, under some assumptions on image statistics and sparsity of coding, the watermarked pixel locations can be estimated for images by using an independent component analysis approach. This latter method relies on the DM signal being independent of the image statistics and so an independent component analysis should isolate the watermark in one of the independent components. This is probably one of the most efficient current attacks. It relies on assumptions of independence of the DM watermark embedding method from the natural image statistics, and hence can be circumvented by making the watermark embedding dependent on the image statistics. Though sufficient literature exists on the importance of key security for lattice based watermarking techniques, except for the work presented in [10], the work is theoritical. We could not find any references related to the investigation of key security applied to time series data.

The work presented by Giakoumaki et. al [33] uses this concept of the secret embedding keys to show that the watermarking method adopted by them for biomedical images is secure. We use the same embedding method to show how the DM-QIM watermark embedding method is insecure.

Messages are embedded in time series data using the embedding method presented in [33]. A method to find the secret embedding keys thus enabling the modification of the watermarked samples is presented. The experiments are based on the well known Kerckhoffs' principle that states that the security of the communication process is based on the secret key. It is assumed that the attacker knows everything about the communication process. In this chapter an efficient distribution-independent approach to attacking watermarks embedded using transform domain based approaches and DM-QIM embedding techniques, using principles from information theory ([61]) and neural networks ([13, 68]) is proposed. It employs a method to estimate the probable location of the hidden information when only a single copy of the watermarked content, an extreme case of the WOA class of attacks, for the discrete wavelet transform (DWT) domain and independent component analysis based DM-QIM watermarking methods. The paper illustrates the fallibility of DM-QIM for time series data. We show that the embedded method used in [33] is insecure.

3.2 Design of a Biomedical Watermarking System

The response of the Human Audio/ Visual System (HA/VS) is usually relied upon for efficient concealment of data (watermark) within other data (cover). Watermark embedding is done such that it is imperceptible to the HA/VS but can be read by systems designed to trace its presence and possibly retrieve it from a given cover work. A robust watermark is one which tends to survive combinations of signal processing attacks while a fragile watermark is highly sensitive to any signal processing technique. In most cases, the robust watermark is used to prove ownership while the fragile watermark is used to prove the extent of the attack. The robust watermark will therefore be embedded in locations of the cover text which carry vital and significant information which result in low data rate (length of the watermark). Any random attack to destroy the robust watermark will automatically destroy the use of the cover work completely. Precise knowledge of the location of the robust watermark in the cover work is hence necessary to destroy it and maintain the quality of the cover work.

A watermarking system is thus designed by an appropriate choice of domain (time, frequency) of the cover work, the transform used and the embedding technique for a given range of attacks (compression, filtering, scaling, cropping, rotation, additive noise) based on the application and use of the embedded information.

Figure 3.1 depicts a two-tier approach to designing a secure watermarking method. Tier-I, the section denoting encryption of the message before embedding is optional. The second part, tier II shows a watermarking system. The watermark and the cover work are input to the watermark embedder for watermark embedding. The watermarked cover is transmitted wherein it is corrupted due to unintentional signal processing methods or intentional attacks. The watermark decoder uses the secret keys to detect the embedded message which could be decrypted (if it was encrypted before embedding). Chapter 2 demonstrated that the DWT, PCA and ICA provide non-interfering channels to embedding multiple watermarks. Since PCA and ICA are both data dependent transforms and use of ICA is advantageous compared to PCA (chapter 2), multiple watermarks embedding using DWT and ICA methods and QIM based embedding techniques are investigated. The design of the multiple watermarks embedding system using DWT or ICA was based

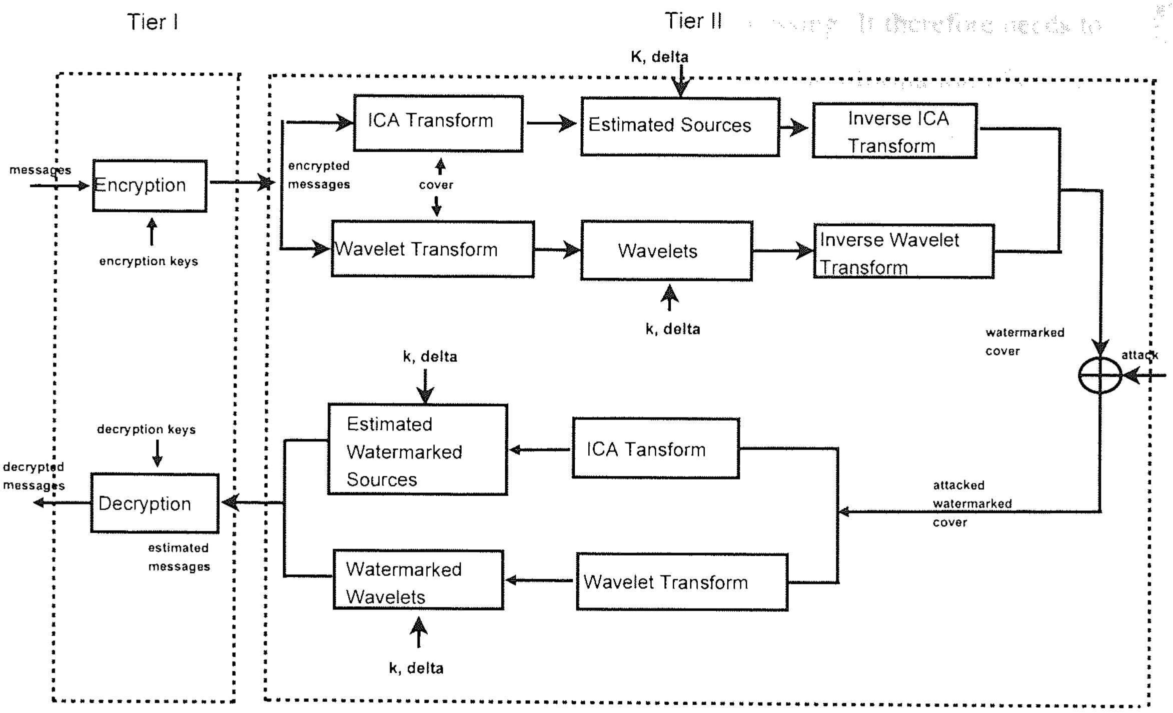


Figure 3.1: Two Tier Approach of Providing Security to Medical Records.

on the trade-off of the three properties of the watermark: robustness, imperceptibility and data rate. Unlike audio/speech watermarking there exists no particular model to test for the distortion to the medical data due to the embedded information. Since the results of medical tests such as EEG, X-Ray images are viewed by the clinician to perform a diagnosis, quantitative measures such as the mean square error discussed in chapter 1 will be used to estimate the distortion. Security of the watermarking system is relied on the secret keys used to embed the watermark. It will be shown that the secret key is vulnerable to estimation attacks and that an attacker could rewrite the entire embedded watermark.

3.2.1 Watermark embedding

The host signal/cover $\mathbf{c}$ is converted from the spatial domain to a transform domain representation $\mathcal{C}$ using a suitable transform denoted by $\mathcal{T}$. Given $\mathbf{c}$ of length $\mathbf{N}_{samp}$ the length of $\mathcal{C}$ is given by $\mathbf{N}_c$ where $\mathbf{N}_c$ is defined by $\mathcal{T}$ and $\mathbf{N}_c \leq \mathbf{N}_{samp}$.

$$\mathcal{T}(\mathbf{c}) \rightarrow \mathcal{C}. \qquad (3.1)$$

The transforms applied to the data include DWT and ICA in this chapter. The watermark embedding is as shown in figure 2.15. A watermark used for copyright purposes requires the highest level of robustness against any further signal processing. It therefore needs to be embedded in selected co-efficients of $\mathcal{C}$ that contain significant information of $\mathbf{c}$. This ensures that the embedded watermark is preserved after any compression of $\tilde{\mathbf{c}}$.

In the DWT method the watermark is embedded in the detail co-efficients obtained for the fourth level of decomposition for the time series data. Since the work presented in this chapter is an analysis of the key security of the watermarking algorithm presented in [33], the choice of wavelets and the watermark embedding technique based on the DWT method derived is the same as presented by Giakoumaki et. al. The authors provide proof showing that the detail co-efficients chosen contain significant energy (information content of $\mathbf{c}$) to survive compression attacks. But these co-efficients do not contain the maximum information content of $\mathbf{c}$ that any distortion to these components results in a large $\mathcal{D}_{Emb}$. The method used to quantify the energy content was tested and the results obtained for the choice of coefficients were similar to that presented by Giakoumaki et. al. More details of the experiments and the results are presented in the results section.

In the ICA method the watermark is embedded in the samples of one of the sources estimated from the cover work (time series) which represents a broadband though informative source of $\mathbf{c}$.

3.2.2 Watermark generation, QIM and Watermark embedding using QIM

The embedded watermark in the following experiments is a binary string of length $\mathbf{N}_{WM} < \mathbf{N}_c$ generated randomly by the owner of $\mathbf{c}$ or the binary representation of textual information. Each bit of this binary string is used to modulate one sample of $\mathcal{C}$. Let $\mathbf{K}$ represent a set of elements of $\mathcal{C}$ which are robust to a defined attack, such as compression. Let $[\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{Nk}] \in \mathbf{K}$ represent the number of different possible sub-sets of length $\mathbf{N}_{WM}$. The WM is embedded in one of the subsets chosen randomly, referred to as $\mathbf{k}$. This vector representing the location of the chosen samples is the secret embedding key.

The QIM method of WM embedding has already been defined in previous chapters. 'QIM' refers to embedding information by first modulating an index or sequence of indices with the embedded information and then quantising the host signal with the associated quantiser or sequence of quantisers [20]. Quantisation index modulation generates the watermarked $\mathcal{C}$, $\tilde{\mathcal{C}}$.

Applying the inverse of $\mathcal{T}$ to $\tilde{\mathcal{C}}$ the watermarked host signal is obtained.

$$\mathcal{T}^{-1}(\tilde{\mathcal{C}}) \rightarrow \tilde{\mathbf{c}}, \tag{3.2}$$

where $\tilde{\mathbf{c}}$ is the watermarked host signal. Dither Modulation (DM) QIM adds a dither signal $\mathbf{o}$ to $\mathcal{C}(\mathbf{k})$ to further secure $\mathbf{m}$. The distortion to $\mathbf{c}$ due to the WM represented as $\mathcal{D}_{Emb}$ is obtained as

$$\mathcal{D}_{Emb} = \mathbf{c} - \tilde{\mathbf{c}}. \tag{3.3}$$

The security of the watermarking system is largely dependent on $\mathbf{k}$ as mentioned earlier, since an estimation of even 75% of $\mathbf{k}$ can lead to loss of copyright as the attacker can overwrite the embedded WM.

All the following experiments are based on the assumption that the attacker knows everything about the system except the keys which are $\mathbf{k}$ and δ. The attacker's challenge is to estimate $\mathbf{k}$ and δ sufficiently close to allow isolation of $\mathbf{m}$.

3.2.3 Data and experiment

The data used is a one-dimensional time series data (single channel EEG sampled at 250Hz as shown in figure 3.2(top)) . The example data (time series) was transformed using both DWT and ICA approaches. The co-efficients obtained after transformation were watermarked by a random binary signal. The watermarked data was tested for robustness and imperceptibility for different values of δ against compression, low pass filtering, noise and varying sampling rates.

The data from the attacked watermarked cover, figure 3.2 (middle), $\hat{\mathbf{c}}$ is used for further experiments on testing the security of the QIM based embedding techniques. figure 3.2 shows the difference between the original EEG and the attacked watermarked EEG, the attack shown is a compression of the EEG signal. It can be seen that the distortion to the original EEG due to the watermark and the attack is minimal. The distortion due to the various attacks has been discussed in chapter 2, hence it has not been repeated. First the security of the DWT method using QIM based embedding will be examined, since, the transform co-efficients in the DWT $\mathcal{T}_{DWT}$ are fixed. $\mathbf{c} \xrightarrow{\mathcal{T}_{DWT}} \mathcal{C}$. $\mathcal{T}_{DWT}$ need not be stored or transmitted to the decoder unlike the ICA method where the exact $\mathbf{W}$ estimated at the embedder is required at the decoder for $\hat{\mathbf{m}} \sim \mathbf{m}$ when η is zero.

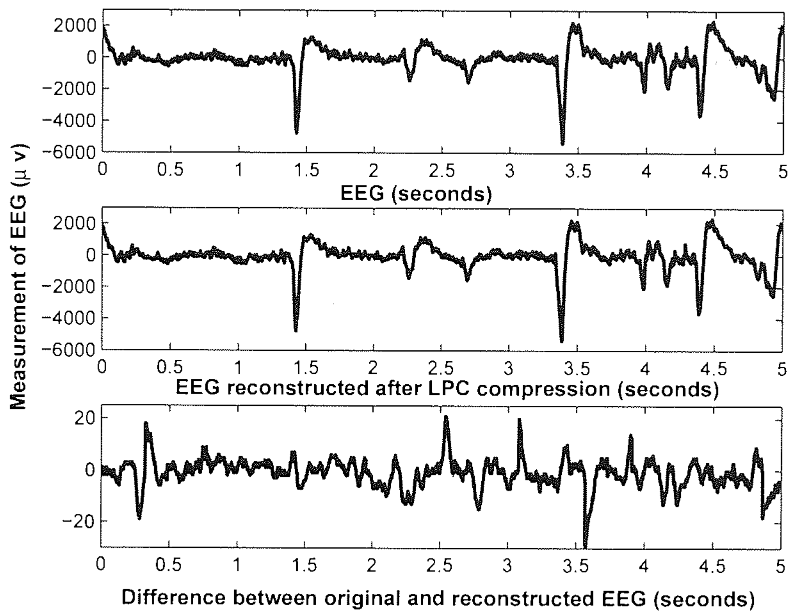


Figure 3.2: Single channel EEG (above), watermarked EEG after compression (below).

3.3 Estimation of the Secret Embedding Key, k

Estimation of the indices of the samples of C used to embed the watermark is dependent on the estimation of δ. This is because the watermarked samples lie on quantiser levels represented by multiples of δ. By estimating δ, the quantiser levels can be generated which in turn help estimate $\mathbf{k}$.

3.3.1 Finding the quantisation index, δ

The robustness provided by the embedding technique based on QIM to the embedded WM is evaluated by subjecting $\tilde{c}$ to different signal processing attacks η. We assume that $\tilde{c}$ is additively modified by η where the modification is represented by ε_c.

$$\tilde{c} + \varepsilon_c \rightarrow \hat{c}. \tag{3.4}$$

The level of degradation of the estimated WM recovered from $\hat{c}$ gives a measure of the effectiveness of a particular embedding technique. While a large δ provides more robustness and increases $\mathcal{D}_{Emb}$ to a higher extent, a small δ decreases $\mathcal{D}_{Emb}$ but makes the WM fragile (figure 3.3).

Figure 3.3 is a plot of the measure of the bit error rate *dist* calculated with reference to the WM retrieved from $\hat{c}$ (time series data) subjected to two different signal processing

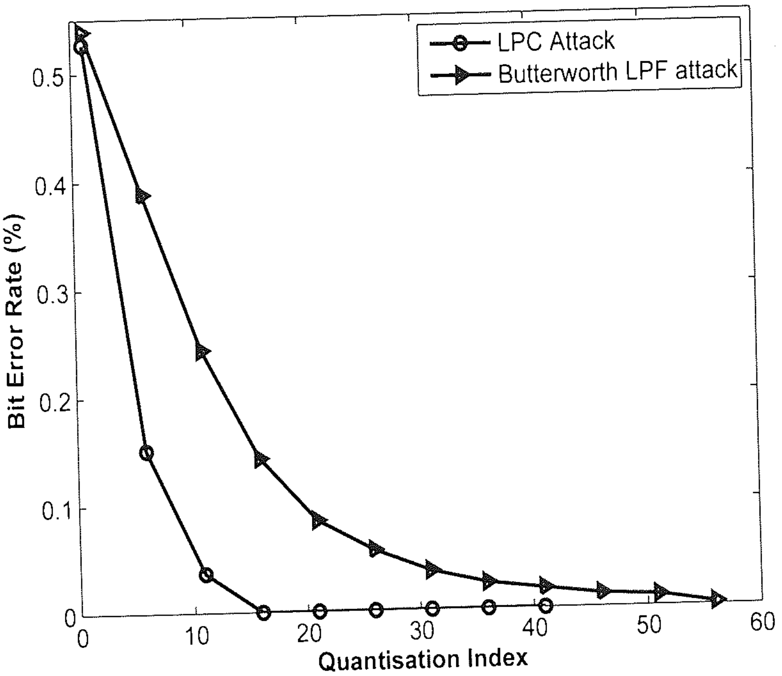


Figure 3.3: Bit error rates of the reconstructed WM for different quantisation indices against two different attacks.

attacks of differing filtering strengths. It can be seen from the results that for the mild attack (LPC compression which allows signal components of most frequencies of $\tilde{\mathbf{c}}$) a smaller value of δ is sufficient to reconstruct the message but for a severe attack (Butterworth low pass filter of order 7, and normalised cut-off frequency 0.75Hz) a larger value of δ is required to obtain $dist = 0$.

Figure 3.4 is the plot of the detail co-efficients obtained for the fourth level DWT decomposition of the one-dimensional EEG signal. Let S be the size of each bin of the histogram and **B** the total number of bins. Then the number of samples of C in each bin represented by b is obtained as follows. So

$$\mathsf{S} = \frac{max(C) - min(C)}{\mathbf{B}} \tag{3.5}$$

and,

$$\mathsf{N}_b = [min(C) + \mathsf{S} * (b-1), \ldots, min(C) + \mathsf{S} * b]. \tag{3.6}$$

where $b \in \{1, \mathsf{N}_b\}$.

The figure 3.4 hence represents the distribution of the elements of C and not the structure of the lattice of quantisers on which the elements of C lie.

Therefore they provide negligible information about the value of δ. Chandrashekar et

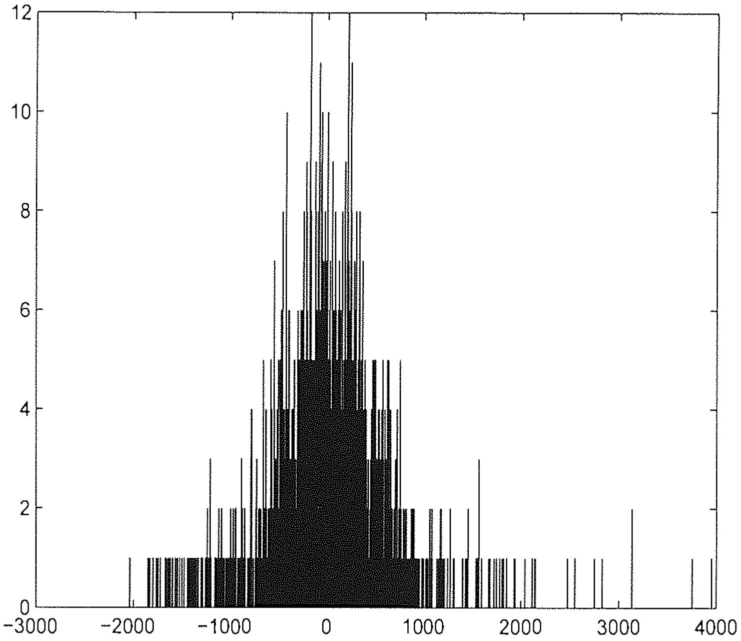


Figure 3.4: Histogram of detail co-efficients of fourth level decomposition of one-dimensional EEG.

al [91] commented that the co-efficients of a transform (they tested the DCT co-efficients obtained for image data) lie centred around zero when they are not watermarked but when they are watermarked they are centred around multiples of δ. By using a low data rate the presence of the WM in a cover work can be disguised. Hiding watermarks by containing the data rate is also discussed in [92]. They surmise that by limiting $\mathcal{R}$ a zero divergence between the $\mathbf{c}$ and $\tilde{\mathbf{c}}$ can be achieved. Distinguishing $\tilde{\mathbf{c}}$ from $\mathbf{c}$ in such a case is not possible. They conclude that for QIM based embedding in Gaussian covers watermarking a third of the coefficients available to carry the watermark, result in zero divergence, for 90% of $\tilde{\mathbf{c}}$. Johnson et al in their book on information hiding [47], also use histograms to detect the presence of embedded messages in images. They use the differences of adjacent histogram values to predict hidden content. It can be seen from these histograms the distribution of $\hat{C}$ and probable clues about hidden information, but not probable values of δ.

Let $\Delta = [d1, d2, \ldots]$ represent the spacing of the lattice on which all the elements of C lie. Watermark embedding using QIM alters the structure of this Δ lattice. $C(\mathbf{k})$ will instead lie on a lattice defined by δ and further randomised by $\mathbf{o}(\mathbf{k})$. Let Γ denote the combined lattice structure of Δ and δ, $\Gamma = [\gamma1, \gamma2, \ldots]$. Since access to $\tilde{\mathbf{c}}$ and not $\mathbf{c}$ is provided, the histogram in figure 3.5 which depicts the representation of C on the γ lattice is derived as follows:

1. $\mathcal{T}(\hat{\mathbf{c}}) \rightarrow \hat{C}$.

2. Sort the samples of $\hat{C}$.

3. $diff(j) = \hat{C}(i+1)$ - $\hat{C}(i)$ where $i = 1$ to length(C) and $j \in \{1, length(C)-1\}$.

4. Plot the histogram of $diff$.

Similar to the histogram in figure 3.4 , let S_γ be the size of each bin of the new histogram shown in figure 3.5. N_{b_γ} represents the number of elements of $diff$ in each bin b_γ.

$$\mathsf{S}_\gamma = \frac{max(diff) - min(diff)}{\mathsf{B}}, \tag{3.7}$$

$$\mathsf{N}_{b_\gamma} = [min(diff) + \mathsf{S}_\gamma * (b-1), \ldots, min(diff) + \mathsf{S}_\gamma * b]. \tag{3.8}$$

The histograms observed in figures 3.4 and 3.5, vary largely because the two sets of plots represent different characteristics of C for the time series data.

The changes in the histogram when $\mathbf{N}_{WM}$ is varied are also distinct. When $\mathbf{N}_{WM}$ is more than 90% of $\mathbf{N}_c$, the histogram resembles an impulse function. This is because $\Gamma \simeq \delta$ and almost 90% of the elements in $diff$ will have the same value. The mean of the elements of $diff$ in this case will be close to δ.

For $\hat{\mathbf{c}}$ to be viable for use, the value of $\mathcal{D}_{Emb} + \varepsilon_{\mathbf{c}}$ must be as small as possible. Since the value of η cannot be controlled, in QIM based watermark embedding the parameter δ defines the position of the watermark on the trade-off triangle. For the embedded watermark to be robust against an attack the value of δ should be large and to comply with the distortion constraint for $\mathcal{D}_{Emb}$, the function $\mathcal{F}_m$ should be close to the identity mapping implying $\delta \simeq \Delta$. The elements of $\tilde{C}$ will therefore lie on the grid defined by $\Gamma = [\gamma_1, \gamma_2, \ldots]$. Therefore the histogram of $diff$ (figure 3.5) can be used to obtain an approximate estimation of δ used at the embedder say $\hat{\delta}$.

Since $\tilde{\mathbf{c}}$ is distorted by a value $\varepsilon_{\mathbf{c}}$ before and, during transmission $\tilde{C}$ is also distorted. Let us represent this distortion as ε_C.

$$\tilde{\mathbf{c}} + \varepsilon_{\mathbf{c}} \Longrightarrow \hat{\mathbf{c}}. \tag{3.9}$$

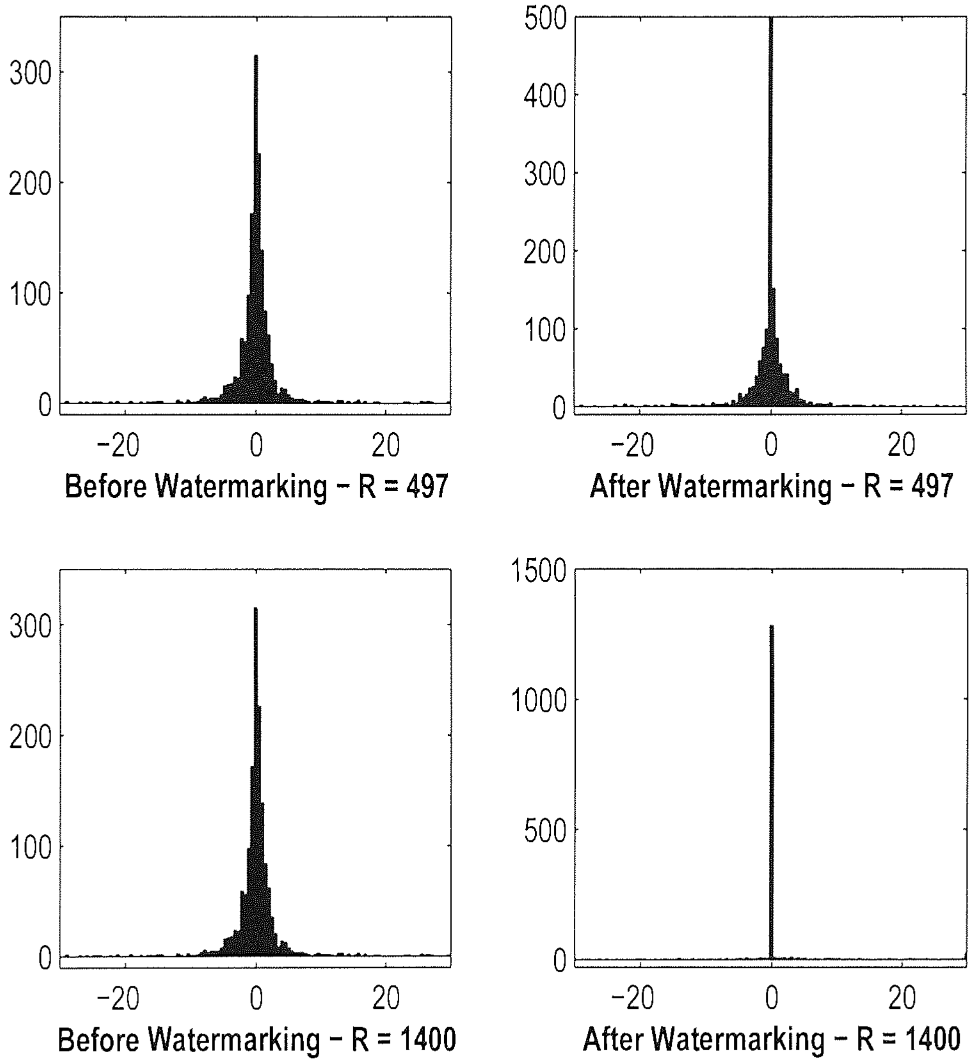


Figure 3.5: Histogram of differences between successive sorted detail co-efficients of fourth level decomposition of one-dimensional EEG for different data rates.

$$\tilde{C}+\varepsilon_C \Longrightarrow \hat{C}. \tag{3.10}$$

$$(\frac{\hat{C}(k)}{\delta})\delta \neq (\frac{\tilde{C}(k)}{\delta})\delta. \tag{3.11}$$

Separating a quantiser level defined by δ from a quantiser level defined by Δ in Γ and calculating the effect of $\mathbf{o}$ (in case of DM-QIM) on a particular $\hat{C}_j$ is not possible. Hence grids of $\breve{\delta}$ on $\tilde{C}$ where $\breve{\delta} = [0,\ldots,\upsilon]$. υ defines the value of $\breve{\delta}$ beyond which the length of $\hat{\mathbf{k}}$ remains constant are applied. Quantifying the value of η and the resulting $\varepsilon_\mathbf{c}$ for each $\tilde{\mathbf{c}}_j$ and consequently $\tilde{C}_j$ is not possible. Hence $\hat{\mathbf{k}}$ is an estimate of the samples of $\hat{C}$ which lie in the range ε of multiples of $\breve{\delta}$. ε is defined to be a fraction of the dynamic range of $\hat{C}$. This is because ε is estimated to be of a value that will not shift $\tilde{C}(\mathbf{k})$ to adjascent (incorrect) quantiser levels. A large ε will estimate a large number of true positives and false positives while a small ε will estimate a small number of true positives and false positives.

$$\breve{\mathbf{k}}_u = [j : \frac{\hat{C}_j}{\breve{\delta}} \leq \varepsilon], \tag{3.12}$$

where u denotes an index $\in \{1, Length(\breve{\delta})\}$ and $j \in \{1, Length(\hat{C})\}$. The length of $\hat{\mathbf{k}}_u$ varies as follows:

1. If $\breve{\delta} \ll \delta \rightarrow$ length of $\hat{\mathbf{k}}_u$ spans the length of $\hat{C}$.

2. If the value of $\breve{\delta}$ increases the length of $\hat{\mathbf{k}}_u$ decreases. This is because the number of quantiser levels decreases.

3. When $\breve{\delta} \simeq \delta$ the length of $\hat{\mathbf{k}}_u$ increases. This is because $\hat{C}(\mathbf{k})$ will lie around the lattice of δ. This is the estimated value of δ represented as $\hat{\delta}$.

4. As the value of $\breve{\delta} > \delta$ the length of $\hat{\mathbf{k}}_u$ decreases.

5. As the value of $\breve{\delta} \gg \delta$ the length of $\hat{\mathbf{k}}_u$ tends to a constant. This value of $\breve{\delta}$ after which the length of $\hat{\mathbf{k}}_u$ does not change defines υ.

This can be seen in figure 3.6 which is a plot of the the estimated $\hat{\mathbf{k}}$ for each value of $\breve{\delta}$.

3.3.2 Estimation of the probable quantisation index, δ

Method I

To estimate the probable value of δ from figure 3.6, applying a baseline correction to the contour of figure 3.6 is necessary. This is to nullify the effect of the value of $\breve{\delta}$ on the

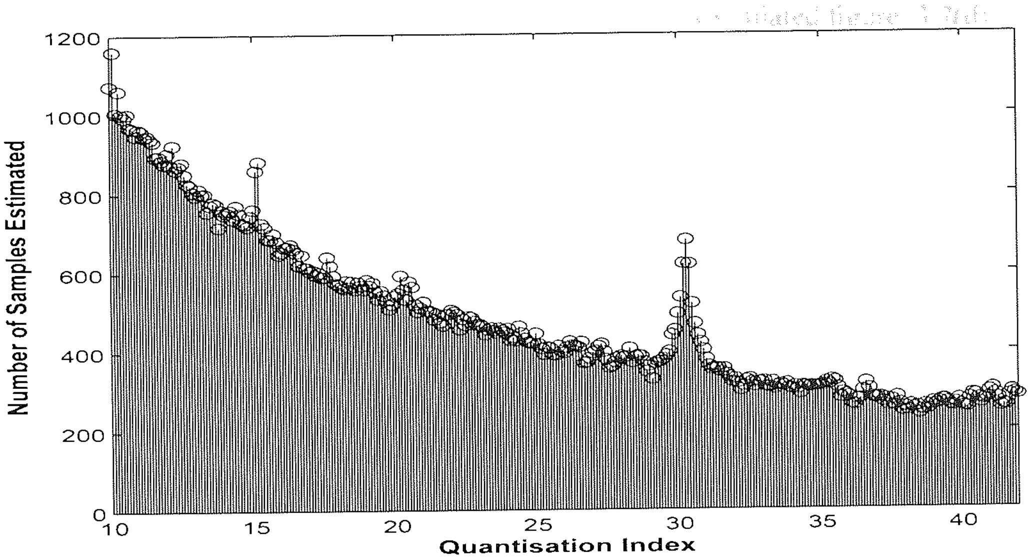


Figure 3.6: $\mathbf{k}$ for different values of $\check{\delta}$.

estimation of $\hat{\mathbf{k}}$ already mentioned. When $\check{\delta}$ approaches the true value of δ, the number of samples of $\hat{C}_u$ satisfying the condition in equation (3.12) increases, thus increasing the length of $\hat{\mathbf{k}}$. As $\check{\delta}$ diverges from the true value of δ, the number of samples of $\hat{C}_u$ satisfying the condition in equation 3.12 decreases, in turn decreasing the length of $\hat{\mathbf{k}}$. But the number of points at which the baseline is corrected will affect the estimation of $\hat{\delta}$. An automated decision on the number of baseline correction points is not possible. Hence the area of each peak is calculated instead. This is done as follows:

- Consider the contour of the plot denoting the estimated $\hat{\mathbf{k}}$ for different values of $\check{\delta}$. The contour of the plot in figure 3.6 decreases with increasing $\check{\delta}$ as explained above with peaks observed for certain values of $\check{\delta}$ Figure 3.7 (a).

- Since there exist multiple peaks, estimating the total number of peaks and the largest peak (which indicates the probable value of δ) among them is required to estimate the value of $\hat{\delta}$. This is done by finding all the minima of the contour plot of figure 3.7 (b). A peak is defined as the largest value of $\hat{\mathbf{k}}$ between two consecutive minima.

- To estimate the largest peak among all the peaks identified, the area between each pair of consecutive minima is calculated. Baseline correction is achieved by interpolating the x-axis representing the distance between the minimal points. That is, straight lines are drawn to connect each pair of consecutive minima as shown in figure 3.7 (c).

- The area between each pair of consecutive minima is calculated figure 3.7(d).

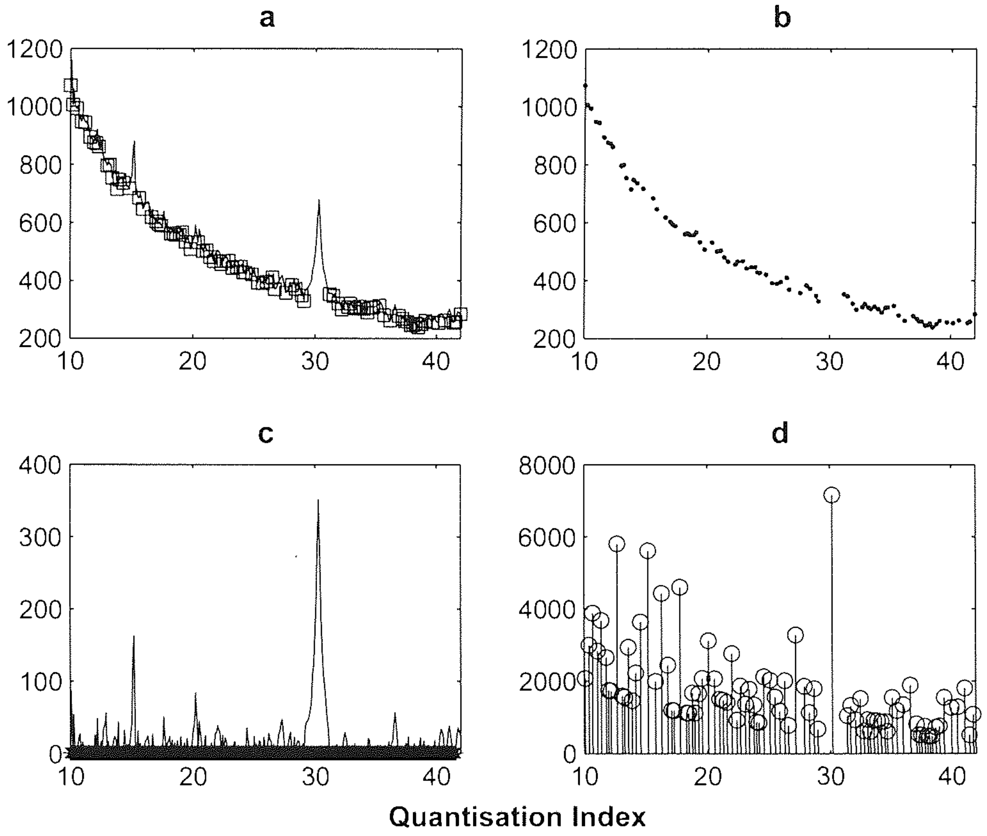


Figure 3.7: Estimation of $\hat{\delta}$ and $\hat{\mathbf{k}}$.

The curve with the highest area denotes the value of $\check{\delta}$ closest to δ, $\hat{\delta}$. From figure 3.7, $\hat{\delta}$ is 30.2. The value of δ used at the embedder is 30.27. The quantised samples were randomised using a dither signal of $\mathcal{N}(0, 1)$.

Method II

To automate the process of detecting the likely value of δ used, we seek a maximum-likelihood-estimator. We use a simple model of a Cauchy-Lorentz distribution for the local distortions induced by the hidden message which is optimised assuming the data samples are i.i.d. ăThe location parameter that maximises the likelihood of the data being represented by a Cauchy distribution indicates the most likely value of δ.

$b(\delta)$ is the plot shown in figure 3.6. $b(\delta)$ arises from the reciprocal nature of the level spacing and the distribution of signal values and distorts the structure due to the hidden message. Therefore in the automated process this baseline effect is first removed before the distribution modelling is performed. We use a thin-plate spline approach using knots determined by locally minimum values so as not to interpolate the structure. Once the background has been removed, the resulting distribution is analysed automatically for a point estimator of the most likely δ.

Specifically, we assume that the likelihood of the sampled data $K = \{k(i), i = 1, \ldots, N\}$ $P(K|\theta)$ given the the parameter set $\theta = (d, \Gamma)$ (in the case of the Cauchy distribution assumed here, this equates to the location parameter d and a halfwidth parameter Γ) is simply $\prod_{i=1}^{N} P(k_i|\theta)$.

Assuming a functional form for $P(k_i|\theta)$ as a Cauchy distribution parameterised by location d and width Γ, gradient optimisation of the likelihood based on the data can be performed to return the most likely parameter choice of location d and hence an estimate of $\hat{\delta}$.

This approach has no prior information on the distribution of the parameters, and so it assumes $P(\theta) = 1$, which is of course generally incorrect since we often know something about the location and the width parameters. If we have some knowledge on these parameters then we can consider estimating the posterior distribution of the model parameters given the data.

$$P(\theta|K) = P(K|\theta)P(\theta)/P(K) \tag{3.13}$$

Since the prior over the data $P(K)$ does not depend on the model, it can be neglected in the optimisation process. So, although we can extend the method to consider a more Bayesian approach to estimating $\hat{\delta}$ if we have additional knowledge on the parameter priors, for the problem considered here across many signal examples we have found that the direct MLE approach provides a simple and robust point estimate of the useful $\hat{\delta}$.

Figure 3.8 shows the result of the estimation of the probable quantisation index $\hat{\delta}$ using the MLE method. The value of $\hat{\delta}$ is estimated by the $\check{\delta}$ at which the peak is obtained.

For the example file shown in Figure 3.6 for illustration, and from Figure 3.8, $\hat{\delta}$ is 30.2. The value of δ used at the embedder for this case was 30.27. The quantised samples were also randomised using a dither signal of $\mathcal{N}(0,1)$. It can be seen that both the estimation methods result in the same value of probable quantisation index.

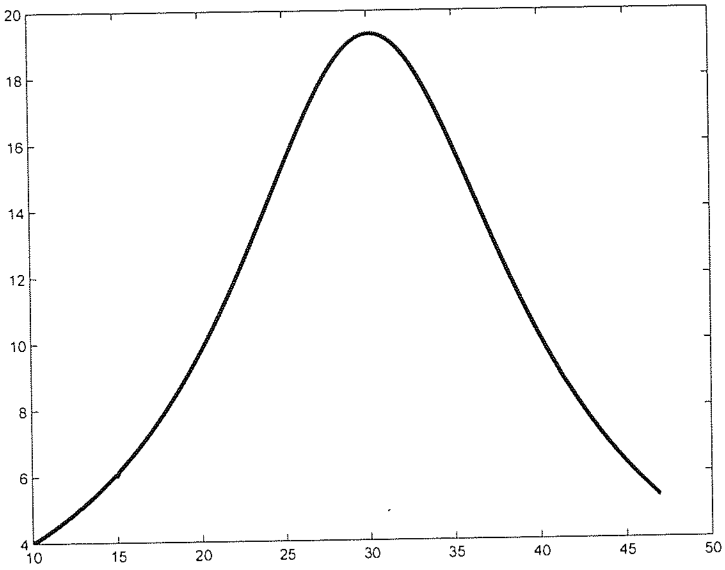


Figure 3.8: Estimation of $\hat{\delta}$ and $\hat{\mathbf{k}}$.

3.4 Estimation of the Secret Key, k, DWT Based Approach

In this section the evidence of how to determine $\mathbf{k}$ is presented based on the theory discussed in section 3.3.

3.4.1 Watermark embedding using DWT

The aim is to estimate $\mathbf{k}$ from $\hat{\mathbf{c}}$, watermarked to contain a single watermark requiring a high level of robustness against a compression attack.

The detail co-efficients of the fourth level decomposition of the DWT applied to the EEG represent $\mathbf{K}$ in this experiment. $\mathbf{k} \in \mathbf{K}$ are used to embed the WM. The DWT transform is applied to $\mathbf{c}$ to obtain the first level decomposition, the approximate co-efficients C_a and the detail co-efficients C_b of $\mathbf{c}$. The application of the wavelet filters to obtain a multiple level decomposition of $\mathbf{c}$ has been discussed in chapter 2.

3.4.2 Watermark detection

The watermarked time series data is subjected to compression attacks as mentioned in section 3.2.3. $\hat{\mathbf{c}}$ obtained in each case is processed to obtain the corresponding $\hat{C}$. The experiments detailed in section 3.3 are conducted on the derived $\hat{C}$ to find $\hat{\delta}$. An estimate of the probable watermarked samples $\hat{\mathbf{k}}$ is the $\check{\mathbf{k}}$ obtained for $\hat{\delta}$.

$$\hat{\mathbf{k}} = [j : \frac{\hat{C}_j}{\hat{\delta}} \leq \varepsilon]. \tag{3.14}$$

3.4.3 Results

The histogram of $\hat{C}$ is obtained for the time series data as explained in section 3.3. The histogram provides the possible range of $\check{\delta}$ to be applied to $\hat{C}$ to estimate δ and $\mathbf{k}$.

Scalar QIM

$\check{\delta}$ is varied from zero to the value representing the significant range of the histogram shown in figure 3.5 for the time series data. Estimation of $\hat{\delta}$ and $\hat{\mathbf{k}}$ is conducted for two different values of ε. $\varepsilon = 0.01 * \frac{max(\hat{C})}{min(\hat{C})}$ and $\varepsilon = 0.02 * \frac{max(\hat{C})}{min(\hat{C})}$. The accuracy of the detection mechanism is tested by the number of samples in the intersection of the two sets $\hat{\mathbf{k}}$ and $\mathbf{k}$.

Figure 3.9(a) is obtained for the watermarked time series data when the range of $\check{\delta}$ used is zero to 42 with an incremental step of 0.01 (the output is shown for $\check{\delta} = 10$ to 42). It can be seen that the output obtained changes for certain values of $\check{\delta}$. When $\check{\delta} \simeq \delta$ there is a sudden peak. Peaks are also observed for $\check{\delta}$ values which have common factors with the δ used to embed the WM. The δ used to embed the WM is 30.27 which has common factors with 10.09, 15.135 and 20.18. It can be observed that at these values of $\check{\delta}$ the characteristics of the plot deviates. The higher the common factor the larger the number of samples in $\hat{\mathbf{k}}$.

The two subplots of figure 3.9 (b) and (c) denote the samples estimated with the $\check{\delta}$ where the largest peak is observed. As mentioned earlier, the light colour bar represents the samples of $\hat{\mathbf{k}}$ and the darker shaded region of the plots, the number of samples obtained for $\hat{\mathbf{k}} \cap \mathbf{k}$. The star represents the length of the embedded WM. Figure 3.9(b) is estimated for $\varepsilon = 0.001 * dynamicrange(\hat{C})$ while figure 3.9(c) represents the estimation for $\varepsilon = 0.002 * dynamicrange(\hat{C})$.

To summarise these experiments on scalar QIM, it was noted that it is possible for an attacker to effectively compensate for the QIM method and overwrite the message.

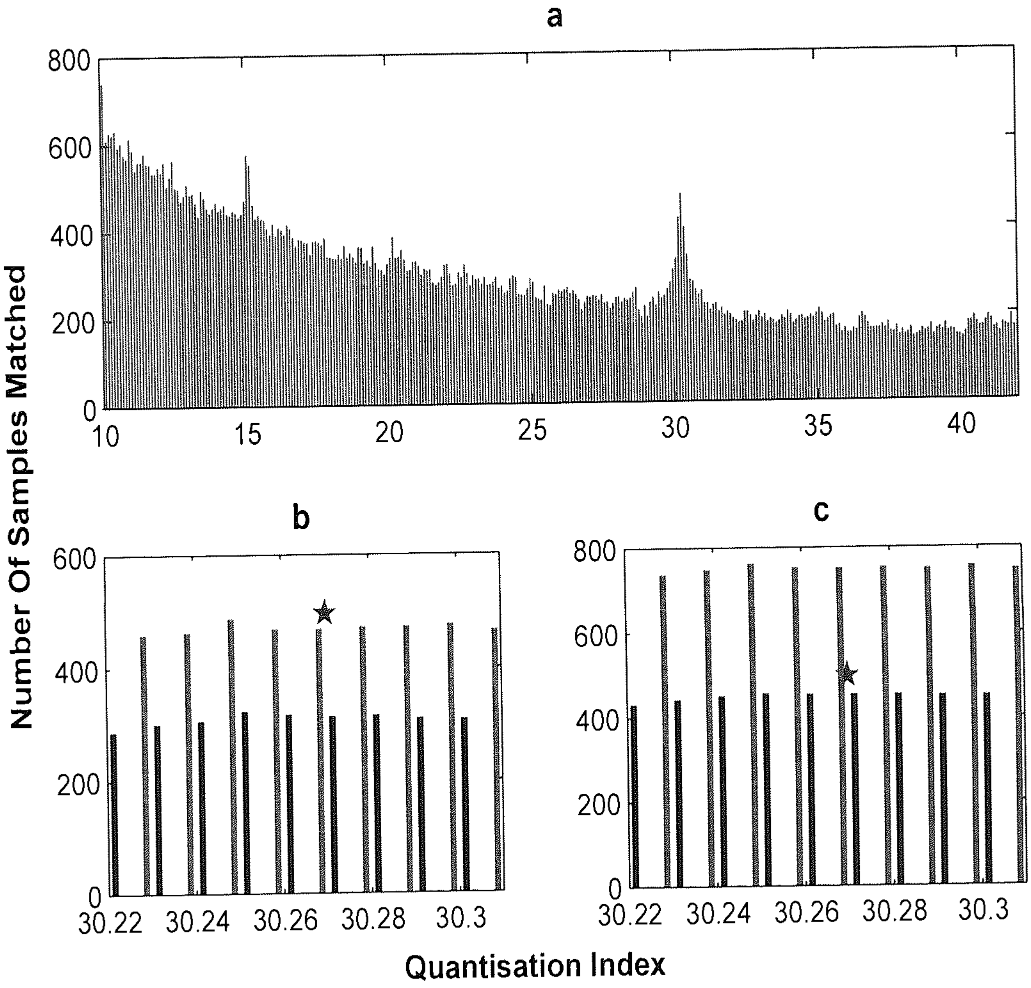


Figure 3.9: a) Estimation of $\hat{\mathbf{k}}$ for time series data using DWT where δ used for *WM* embedding = 30.27 and ε is 0.001 times the dynamic range of $\tilde{C}$. b) Estimated samples of $\hat{\mathbf{k}}$ for $\hat{\delta}$ around the largest peak detected (light colour) and $\hat{\mathbf{k}} \cap \mathbf{k}$ (darker colour), ε is 0.001 times the dynamic range of $\tilde{C}$. The star represents the number of true watermarked samples. c) Estimated samples of $\hat{\mathbf{k}}$ for $\hat{\delta}$ around the largest peak detected (light colour) and $\hat{\mathbf{k}} \cap \mathbf{k}$ (darker colour), ε is 0.002 the dynamic range of $\tilde{C}$. The star represents the number of true watermarked samples.

DM-QIM

In DM-QIM, the security is still dependent on the two factors **k** and δ.

$$(\frac{C_j}{\delta})\delta + o_j \rightarrow \tilde{C}_j \tag{3.15}$$

where j spans the length of **k** and o is a random sequence of values of $\mathcal{N}(0,\sigma)$ distribution. The watermarked samples of $\tilde{C}$ are a multiple of δ shifted by o. Estimation of $\hat{\mathbf{k}}$ in the case of DM-QIM therefore is similar to the estimation in the case of scalar QIM and follows the principle detailed in section 3.3.

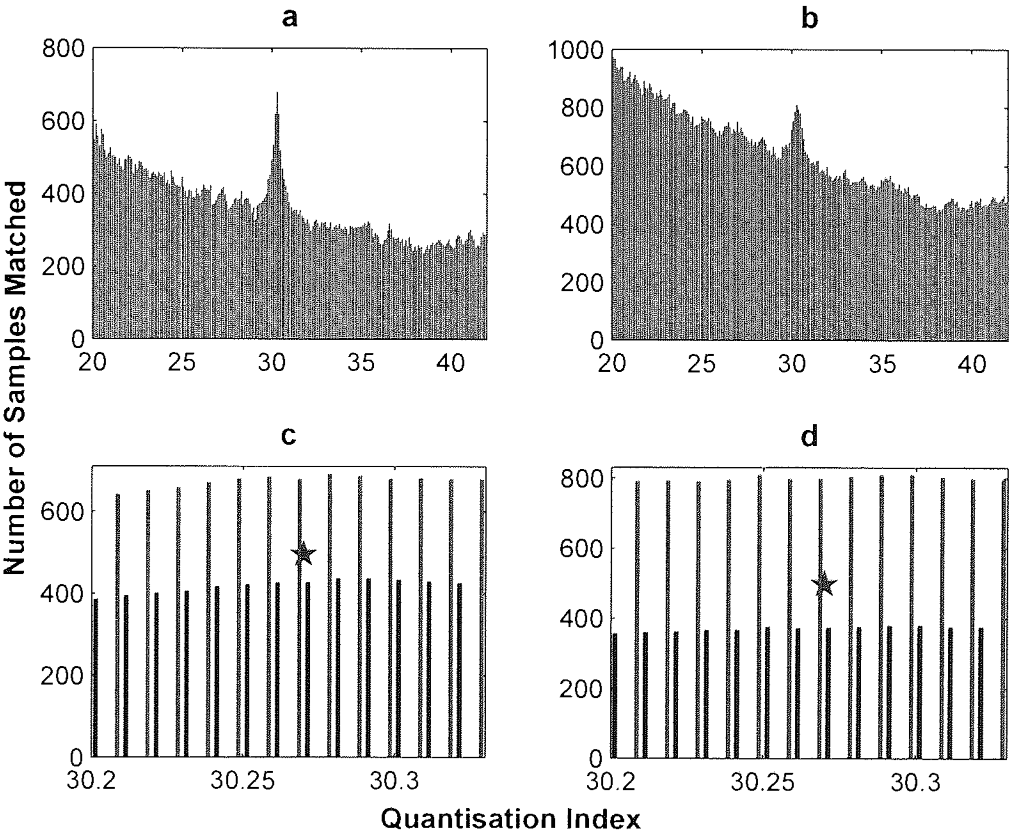


Figure 3.10: a) Estimation of **k** for time series data using DWT where δ used for WM embedding = 30.27 and randomised by adding a dither signal of $\mathcal{N}(0,1)$. b) Estimation of **k** for time series data using DWT where δ used for WM embedding = 30.27 and randomised by adding a dither signal of $\mathcal{N}(0,5)$. c) Estimated samples of $\hat{\mathbf{k}}$ for $\hat{\delta}$ around the largest peak detected (light colour) and $\hat{\mathbf{k}} \cap \mathbf{k}$ (darker colour), the dither signal used to embed the WM is $\mathcal{N}(0,1)$. The star represents the number of true watermarked samples. d) Estimated samples of $\hat{\mathbf{k}}$ for $\hat{\delta}$ around the largest peak detected (light colour) and $\hat{\mathbf{k}} \cap \mathbf{k}$ (darker colour), the dither signal used to embed the WM is $\mathcal{N}(0,5)$. The star represents the number of true watermarked samples.

Figure 3.10(a) denote the samples of C which have been actually watermarked (dark shade) and the estimate of the samples for different values of $\check{\delta}$ (light shade) using DM-QIM. The samples of $\mathcal{C}$ were quantised using $\delta = 30.27$ and a dither signal $\mathcal{N}(0,1)$ at the encoder. Unlike the detection results in the case of scalar QIM, a large number of samples for most values of $\check{\delta}$ were estimated. But as can be seen from figure 3.10(a), the distribution of the estimated samples changes when $\check{\delta}$ approaches the true value of δ as in figure 3.9. The detection method described in section 3.3 is applied to obtain $\hat{\delta}$ and $\hat{\mathbf{k}}$. By concentrating on the values around this $\hat{\delta}$ value, the results shown in figure 3.10(c) are obtained. When $\hat{\delta} \simeq \delta$ most of the samples of $\mathbf{k}$ were estimated and the number of false positives is less than one third the value of $\mathcal{R}$.

The figure 3.10(d) is obtained when $\mathcal{C}$ is watermarked using a dither signal with a large variance. The dither signal used is $\mathcal{N}(0,5)$. This large variance dither signals showed small changes in the structure (figure 3.10(b)). The detection mechanism identified the maximum peak at 30.3 for the time series data. Dither signals with a variance almost equal to the δ value will not be used in practice as it has been observed that they increase the distortion of the cover data significantly. However as a test the WM for this high variance was embedded. The number of false positives is quite high but all the values of $\mathbf{k}$ are estimated correctly.

DM-QIM offers a slightly better security compared to scalar QIM but as can be seen both the embedding techniques are insecure. It is possible to overwrite the embedded message to a large extent without destroying $\boldsymbol{c}$.

3.5 Estimation of the Secret Key, k, ICA Based Approach

The Independent Component Analysis (ICA) estimates underlying sources from a set of mixed observations [42]. Given p observation vectors

$$\mathrm{X} = [\mathrm{x}_1 \mathrm{x}_2 \mathrm{x}_3 \ldots \mathrm{x}_p]' \tag{3.16}$$

the ICA estimates l statistically independent source vectors

$$\mathrm{S} = [\mathrm{s}_1 \mathrm{s}_2 \mathrm{s}_3 \ldots \mathrm{s}_l]' \tag{3.17}$$

and the mixing matrix $\mathbf{A}$ where $l \leq p$,

$$\mathrm{X} = \mathrm{AS}. \tag{3.18}$$

The security of the embedding method in the ICA transform domain is now presented.

The application of the ICA algorithm to a single channel EEG signal to obtain an estimate of the underlying sources has been discussed in chapter 2. The delay embedding method is used to construct a matrix from a single channel EEG explained in chapter 2 for experiments using the ICA method in this chapter.

Some concepts explaining the construction of the input matrix to the ICA are repeated here. The one-dimensional EEG is recorded from an alert adult. It is sampled at 250Hz and each sample is represented as a 16 bit unsigned integer. The slowest signal component is assumed to be 3Hz and a delay of 1 sample between two successive vectors is used. The required required embedding window thus calculated **EmbWin** is 83. **EmbWin** sets the upper bound on the total number of sources p that can be estimated from the EEG. The number of delay vectors N_{ov} for a one-dimensional signal $\mathbf{c}$ of length N_c is given as

$$N_{ov} = \mathsf{N}_c - \mathbf{EmbWin} + 1. \tag{3.19}$$

The matrix $\mathbf{X}$ is derived from the single channel $\mathbf{c}$ as follows:
delay embedding of $\mathbf{c}$ with delay of one sample between two successive delay vectors, $\mathbf{d}_i$ and $\mathbf{d}_{i+1}$, $\mathbf{EmbWin} = p$

$$\mathbf{d}_1 = \mathbf{c}(1,\ldots,p). \tag{3.20}$$

$$\mathbf{d}_i = \mathbf{c}(i,\ldots,p+i-1), \tag{3.21}$$

where $i = 2,\ldots,N_{ov}$.

$$\mathbf{X} = [\mathbf{d}_1;\mathbf{d}_2;\ldots;\mathbf{d}_{N_{ov}}]'. \tag{3.22}$$

The input to the ICA is $\mathbf{X}_{p\times N_{ov}}$

$$[\mathrm{S},\mathrm{W}] \overset{ICA}{\leftarrow} \mathbf{X}. \tag{3.23}$$

As already seen with the experiments on the synthetic data in the previous chapter, the delay embedding method of estimating signals from a one dimensional signal has been shown to extract good estimates of underlying sources. As the input to the ICA contains vectors delayed by one sample N_{WM} is severely constrained. Only one sample across each diagonal of the matrix representing all the estimated sources can be watermarked.

Watermarks embedded in low-middle frequencies of $\mathbf{c}$ are robust to compression attacks. Hence the samples of one the sources, $\mathbf{s}_{wm}$ obtained from the ICA transform representing the low frequency component of $\mathbf{c}$ represent $\mathbf{K}$. A random selection of samples representing $\mathbf{k} \in \mathbf{K}$ is chosen equal to the length of the watermark to be quantised.

$$\mathcal{F}_m(\mathbf{s}_{wm}(\mathbf{k}), WM) \rightarrow \tilde{\mathbf{s}}_{wm}. \tag{3.24}$$

$\mathbf{S}$ is modified to $\tilde{\mathbf{S}}$ in the sense that one of the sources (rows) represented by $\mathbf{s}_{wm}$ is watermarked. The watermarked matrix representing the set of observations is obtained by multiplying $\tilde{\mathbf{S}}$ by the inverse of $\mathbf{W}$, $\mathbf{A}$.

$$\tilde{\mathbf{X}} = \mathbf{A} * \tilde{\mathbf{S}}. \tag{3.25}$$

3.5.1 Construction of the one-dimensional watermarked cover from the watermarked data matrix

For the time series data a one-dimensional $\tilde{\mathbf{c}}$ needs to be reconstructed from $\tilde{\mathbf{X}}$. $\tilde{\mathbf{X}}$ is a matrix whose diagonals from top right to bottom left contain the same element except for certain values which are altered due the embedded watermark. Hence careful re-ordering of the samples is required to prevent loss of the embedded information at the embedding stage itself. This has been discussed in chapter 2. Method 2 of the reconstruction of the one-dimensional $\tilde{\mathbf{c}}$ from $\tilde{\mathbf{X}}$ discussed in chapter 2 is adopted.

3.5.2 Estimation of the secret key, k and the probable quantisation index, δ

The watermarked EEG data $\tilde{\mathbf{c}}$ is subjected to a compression attack η.

$$\tilde{\mathbf{c}} + \eta \rightarrow \hat{\mathbf{c}}. \tag{3.26}$$

In order to estimate the probable δ and $\mathbf{k}$ an estimate of the probable sources will need to be obtained. Hence $\hat{\mathbf{X}}$ is constructed from $\hat{\mathbf{c}}$ as in the watermark embedding process.

Two different experiments were conducted. One, given the separating matrix $\mathbf{W}$ used at the embedder to transform the EEG to the estimated sources, is it possible to find the watermarked source and, δ and $\mathbf{k}$? Experiment two answers the question 'How close can an attacker get to recovering the message, given she knows the method and key factors such as segmentation blocking?'.

Estimation of the secret key, k, sources estimated using the basis vectors used at the embedder

An estimate of the probable sources is obtained as the projection of $\hat{\mathbf{X}}$ onto the independent components, rows of $\mathbf{W}$.

$$\hat{\mathbf{S}} = \mathbf{W}\hat{\mathbf{X}}. \tag{3.27}$$

The histogram for the time series data is shown in figure 3.11. This histogram is obtained as explained in section 3.3 and provides a possible range for $\check{\delta}$. The source whose histogram is plotted is chosen randomly but the histograms of all the sources obtained, as explained in section 3.3, are nearly the same. The estimation of $\hat{\mathbf{k}}$ and $\hat{\delta}$ was conducted similar to that explained in the section 3.3.

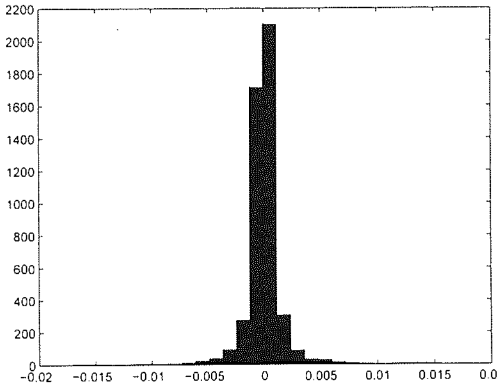


Figure 3.11: Histogram of difference between sorted samples of the estimated source, time series data.

Figure 3.12 is the output of the estimation of $\hat{\mathbf{k}}$ for the watermarked source (time series). It is assumed that the attacker has no knowledge of the identity of $\hat{\mathbf{s}}_{wm}$ in the set of p sources. Therefore the estimation of $\hat{\delta}$ and $\hat{\mathbf{k}}$ is conducted on all the estimated sources. $\check{\delta}$ takes values from zero to 0.2.

The estimation method described in section 3.3 is applied to find the probable values of δ and $\mathbf{k}$. A change in the envelope of the output at a particular value of $\check{\delta}$ (figure 3.12, upper subplot) was noticeable. Further experiments concentrating around the peak (figure 3.12, lower plot) show that $\hat{\mathbf{k}}$ estimated for $\check{\delta}$ contain a large number of samples corresponding to $\mathbf{k}$.

Figure 3.13 is the estimation result obtained for $\mathbf{s} \neq \mathbf{s}_{wm}$. Unlike the results obtained for $\mathbf{s}_{wm}$, there is no distinctive peak and the output is varying randomly. Similar outputs to figure 3.13 were observed $\forall \mathbf{s}_i; i \neq wm$. Also the length of $\hat{\mathbf{k}}_u$ estimated for different $\check{\delta}$

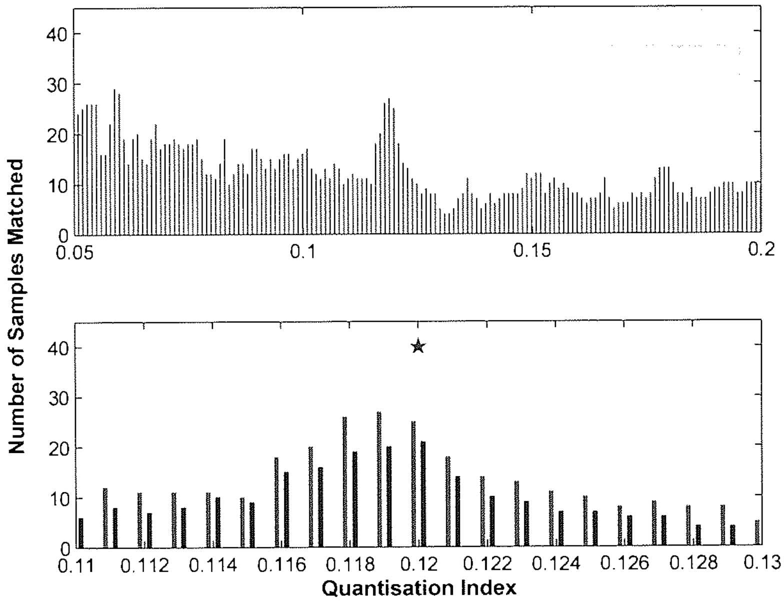


Figure 3.12: Above: Estimation of $\mathbf{k}$ for ICA where $\delta = 0.12$ is used for WM embedding and the sources are obtained from the watermarked EEG using the same separating matrix used by the embedder. The result shown is for the source which has been watermarked. Below: Estimated samples of $\hat{\mathbf{k}}$ for $\check{\delta}$ around the largest peak detected (light colour) and $\hat{\mathbf{k}} \cap \mathbf{k}$ (darker colour). The star represents the true length of the watermarked samples.

in the case of $\mathbf{s} \neq \mathbf{s}_{wm}$ is smaller compared with $\hat{\mathbf{k}}_u$ estimated for different $\check{\delta}$ in the case of $\mathbf{s}_{wm}$. $\hat{\mathbf{k}}_u$ estimated for different $\check{\delta}$ in the case of $\mathbf{s}_{wm}$ also has a distinctive peak compared to $\hat{\mathbf{k}}_u$ estimated for different $\check{\delta}$ for the unwatermarked sources.

Estimation of the secret key, k, sources estimated by applying ICA to the attacked watermarked cover

The sources in this experiment were derived by applying the ICA to $\hat{\mathbf{X}}$. The independent components (the separating matrix $\mathbf{W}$) are derived from the input data. The input at this stage is $\hat{\mathbf{X}}$ the noise contaminated version of $\mathbf{X}$. Hence the independent components obtained from this input are denoted by $\hat{\mathbf{W}}$. Let the sources estimated as projections of $\hat{\mathbf{X}}$ on these independent components be represented as $\hat{\mathbf{S}}_{att}$.

$$ICA(\hat{\mathbf{X}}) \rightarrow [\hat{\mathbf{S}}_{att}, \hat{\mathbf{W}}], \tag{3.28}$$

$$\hat{\mathbf{S}}_{att} = [\hat{\mathbf{s}}_{1_{att}}; \hat{\mathbf{s}}_{2_{att}}; \ldots; \hat{\mathbf{s}}_{wm_{att}}; \ldots; \hat{\mathbf{s}}_{l_{att}}]'. \tag{3.29}$$

The estimation of $\hat{\delta}$ and $\hat{\mathbf{k}}$ is conducted as explained in section 3.3 and the results are shown in figure 3.14. All the estimated sources (rows of $\hat{\mathbf{S}}_{att}$) are sampled for different

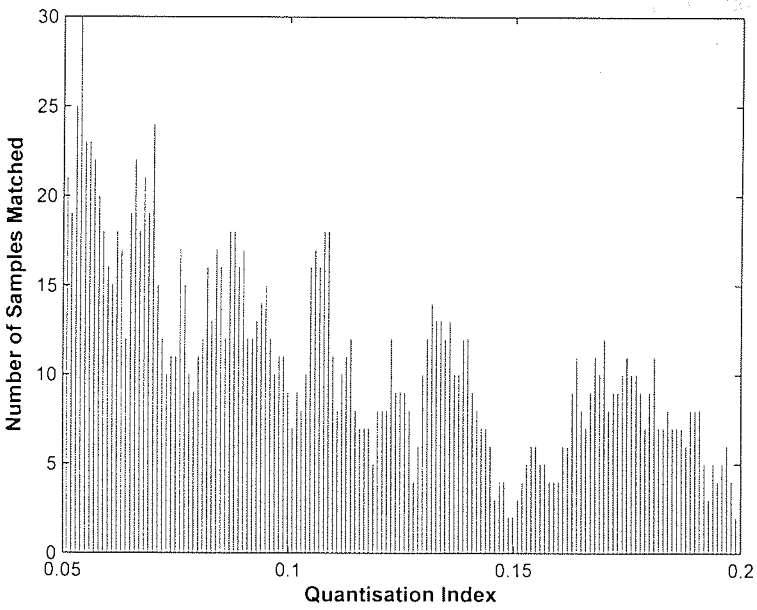


Figure 3.13: Estimation of **k** for ICA where δ = 0.12 is used for *WM* embedding and the sources are obtained from the watermarked EEG using the same separating matrix used by the embedder. The result shown is for the source which is not watermarked.

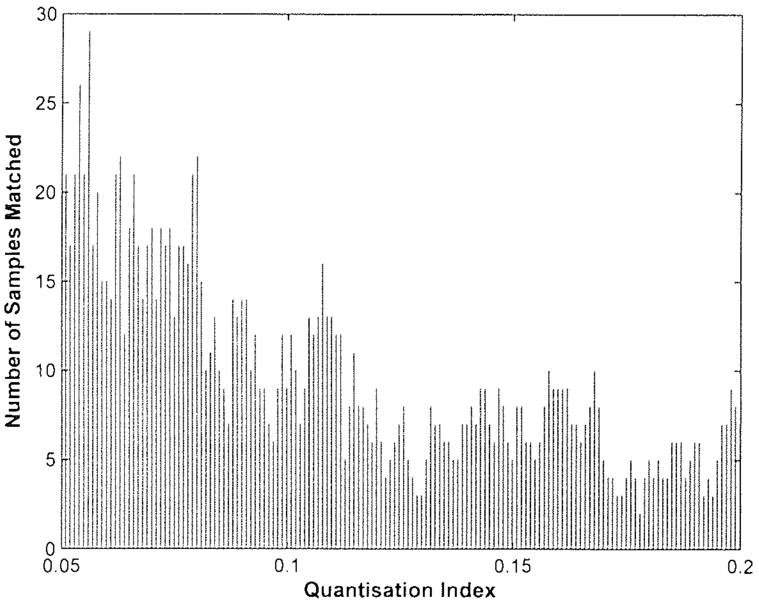


Figure 3.14: Estimation of **k** for ICA method where δ = 0.12 is used for *WM* embedding and the sources are obtained using the separating matrix estimated from the compressed watermarked EEG. The result shown is for a source chosen randomly as the result for all the sources was similar.

values of $\check{\delta}$ starting from zero to 0.2. Figure 3.14 is the result of the estimation mechanism for $\hat{\mathbf{k}}$ for one of the sources estimated by applying the ICA to the EEG data.

It can be seen that no value of $\check{\delta}$ positively identifies the actual δ used at the embedder. This result was observed for all the sources estimated. We conclude that an attacker applying the ICA to $\hat{\mathbf{c}}$ will not be able to estimate $\mathbf{k}$ and hence will be unable to destroy the embedded WM significantly without destroying the cover.

3.6 Summary of the Results

The estimation of $\hat{\delta}$ and $\hat{\mathbf{k}}$ was conducted for 45 different EEG signals taken from different data sets (no epileptic activity, epileptic activity). Each one-dimensional signal was watermarked using a different value of δ and $\mathbf{k}$ using both DWT and ICA based approaches. The watermarked EEG signals were applied various attacks such as differing encoding rates, low pass filtering, addition of Gaussian noise and compression. The results obtained are summarised in Table 1. Correct detection refers to the result obtained when the peak obtained for $\hat{\mathbf{k}}$ was distinctive. Partial detection is the result when more than one peak was estimated. The estimated $\hat{\mathbf{k}}$ for the true value of δ (used at the embedder) contained almost 50% of the total number of samples used to embed the watermark while the second largest peak contained 50% of the total number of samples used to embed the watermark. When no peaks were identified the result is termed as, No detection.

	Correct detection	Partial detection	No detection
DWT	45	0	0
ICA	38	2	5

Table 3.1: Result of estimation of $\hat{\mathbf{k}}$ and $\hat{\delta}$.

Note that our approach is based on a pdf estimation of *differenced* watermarked transform domain, and so enjoys the benefits of being algorithmically simple and not reliant on assumptions about the distributions of covertext or watermark values.

It suffers when pdf estimation of high dimensional multivariate distributions is needed. However, for most applications this is limited to one or two dimensional pdfs and so is computationally tractable. Our method failed when the strength of the attack applied to the watermarked signal resulted in a bit error rate of 30%.

3.7 Conclusion

In this chapter the claim that QIM and DM-QIM are secure embedding techniques were examined by investigating the likely estimation of the unknown δ and $\mathbf{k}$. The experiments were conducted for two transform domain methods DWT and ICA. The results obtained show that the QIM method of embedding is not generally secure as has been claimed. Though retrieval of the exact message is not possible, deleting the existing message or embedding another message which destroys the original *WM* is possible. The results using the ICA method are not as precise as the results obtained for the DWT method. We believe this is due to the way the two transforms derive their basis vectors. In the case of the DWT the transformation of $\mathbf{c}$ to wavelets is obtained by applying an averaging filter having a small number of co-efficients whereas in the case of ICA it involves matrix multiplication and hence minor changes to $\tilde{\mathbf{c}}$ result in completely obscuring the distortion to $\mathbf{c}$ due to watermark embedding using quantisation.

It was verified that the use of ICA for watermarking has the benefit of an in built sensitivity to data snooping. This chapter was an attempt to verify the security of one of the most commonly used watermark embedding techniques. Scalar QIM is shown to be the least secure and DM-QIM performs only slightly better in comparison. It was also shown that the assumption 'QIM based embedding techniques are secure due the random selection of $\mathbf{k}$ and δ' is not generally correct. Using QIM and DM-QIM in the wavelet transform domain still allows a reasonable estimation of δ and $\mathbf{k}$ permitting overwriting of the embedded message without overly increasing the distortion of the cover. The results obtained for the ICA method show that the estimation of $\mathbf{k}$ is more complicated due to the sensitivity of the ICA. The ICA algorithm being data dependent, slight changes to the input data resulted in changed transform co-efficients and estimated sources. At this stage we conclude that in QIM based embedding methods, the use of the dither signal does not provide adequate security. Without a pragmatic approach to information hiding and privacy protection, the widespread deployment of the EPHR is likely to be compromised due to lack of public acceptance.

This chapter verified the security of the DM-QIM based watermark embedding techniques illustrated on time series data. As we have already verified that the ICA method is capable of providing multiple channels to embed multiple hierarchical *WM*s and the sensitivity of the ICA provides an inbuilt security, in the following chapters we will study if the ICA method provides a mechanism to embed multiple watermarks and find the order of the embedded watermarks at the decoder. The results summarised in this chapter were presented in [60].

4 MULTIPLE WATERMARK EMBEDDING

CONTENTS

In the previous chapters the different transform domain steganographic methods including DFT, DCT, DWT, PCA and ICA were discussed. It was noted that the ICA method has an in built security mechanism in chapter 3. It was noted that only three of the transforms DWT, PCA and ICA provide a convenient mechanism to embed multiple watermarks which can be recovered without error at the decoder. Multiple watermarks can also be embedded in the the DFT and DCT domains provided the samples of **k** chosen for each watermark do not overlap. This is because the output of the DFT and DCT transform of a one-dimensional signal is also a one-dimensional signal and both the input and output are of equal lengths.

In the case of the DWT method, the output signal is of the same length as the input signal but by successively applying the DWT transform to the low frequency components obtained at each level the input signal can be further decomposed into more detailed and coarse representations. In PCA and ICA methods, the input to the transform is always a matrix and the one-dimensional signal is transformed into a matrix using the non-overlapping segments or the delay embedding method. The output of these transforms is also a matrix where each vector gives a different representation of the signal **c** with differing characteristics. Since PCA uses the second order characteristics of the input signal and output signals are merely decorrelated, compared to the ICA which estimates signals which are independent. The PCA method provides multiple non-orthogonal channels to embed multiple watermarks. Similarly the DWT method also provides multiple orthogonal channels to embed multiple watermarks. In this chapter the DWT (an example of a non-orthogonal transform) and the ICA (transform providing statistically independent channels) will be evaluated for multiple watermark embedding to assess the method which can provide a better trade-off of the principal characteristics of a watermark compared to watermarking in the spatial domain.

4.1 Multiple Watermarks

Embedding multiple watermarks has been actively investigated in papers from Mintzer and Braudaway's 'If one watermark is good, are more better?' [66] to Sencar and Memon's 'selective detection of embedded watermarks to confuse the attacker' [87]. Current embedding methods have been proposed for multimedia data and in the biomedical domain to medical images. Tao et al [94] discuss the embedding of multiple watermarks in the

DWT domain for image data. They show that the robustness of the watermarks embedded in different levels of the wavelet decomposition of an image is varied based on the type of attack. Wong et al [99] propose a method of embedding multiple watermarks into a cover containing multiple watermarks. The new watermarks are embedded in locations orthogonal to the watermarked locations. Jin et al [46] present a multiple watermarking method using both the DWT and ICA. The multiple watermarks are mixed and demixed using the ICA and are embedded in the wavelet transform domain of **c**. However as seen in the previous chapter very little work exists on watermarking time series biomedical data. A more detailed study of the two techniques (DWT and ICA) for such single channel time series data (EEG and ECG) which have lower redundancy than images is important to derive a data independent, secure standard for an eHealth system. In this chapter a comparison the effectiveness of the DWT and ICA approaches in providing authentication and security to such time series data for the embedding of multiple watermarks. In [62], two visual watermarks are embedded in the DWT domain through modification of both low and high frequency coefficients. Watermark data inserted into low frequencies is more robust to image distortions that have low pass characteristics like filtering, lossy compression, and geometric manipulations but less robust to changes of the histogram such as contrast/brightness adjustment, gamma correction, and cropping. On the other hand, watermark data inserted into middle and high frequencies is typically less robust to low-pass filtering, lossy compression, and small geometric deformations of the image but extremely robust with respect to noise adding, and nonlinear deformations of the gray scale. Since the advantages and disadvantages of low and middle-to-high frequency watermarks are complementary, embedding multiple watermarks in an image (namely, one in lower frequencies and the other in higher frequencies) would result in a scheme that is highly robust with respect to a large spectrum of image processing operations.

Sheppard et al [88] present three different algorithms for embedding multiple watermarks, Re-watermarking, Segmented watermarking and Composite watermarking. Re-watermarking method of embedding multiple watermarks is defined as adding watermarks one after another to a watermarked cover. Segmented watermarking is obtained when the cover work is divided into segments. Each segment is used to carry one watermark. This method of watermarking is necessary when embedding multiple watermarks of dissimilar characteristics into a one-dimensional **c** and the transform applied derives a one-dimensional C. The DFT and DCT transforms are examples of this method as mentioned previously. Composite watermarking is defined as the method wherein a single composite watermark is derived by combining multiple watermarks. The composite watermark is embedded in **c**. The embedding of multiple watermarks in our experiments in the DWT/ICA transforms of **c** does not use any of the three algorithms. In this thesis the multiple watermarks are embedded simultaneously in **c** in multiple non-orthogonal/independent signal components derived from **c**.

4.2 Characteristics of Each of the Multiple Watermarks

Multiple messages (text, audio, image) or different types of **m** are generated, to fulfil the different requirements of security and authentication. These include:

- Tamper Detection, $\mathbf{m}_1$: A random binary string embedded randomly across the entire length of c to provide tamper detection. This has to be known by the decoder at the receiving end.
- Diagnostic Information, $\mathbf{m}_2$: Textual details including specification of medical tests, diagnosis results, doctors' notes. $\mathbf{m}_2$ may be used at the decoder for further processing of the medical data.
- Data Provider Identification, $\mathbf{m}_3$: Clinician's name or identification to be used as record of source identification.
- Patient Privacy, $\mathbf{m}_4$: Patient's personal details embedded to prevent identification of the ownership of a medical record by unauthorised users of the patient data.

As $\mathbf{m}_1$ is a binary string, the derived messages $\mathbf{m}_i$, i = 2 to 4 are converted to binary format WM_i. .

$$\mathbf{m}_i \rightarrow WM_i. \tag{4.1}$$

The WM_4 representing personal details requires a high level of robustness to remain hidden/unaltered from malicious attacks, while WM_1 is used to estimate the level of attack, and hence needs to be fragile and be destroyed when the cover work **c** undergoes any transformation. Hence our requirements list reflects a hierarchy from fragile to robust watermarking.

The different watermarks are embedded in the transformed domain of the EEG **c** using a suitable embedding technique (for example QIM). The DWT and ICA transforms will be used to convert the EEG to wavelets and independent sources respectively.

4.3 Data and Experiment

The single channel EEG considered for the experiments in this chapter is one of a 22 channel EEG recording sampled at 250Hz, and each sample is encoded using signed 16 bit integer format. A one hour recording of this 22 channel EEG needs a storage space greater than 1Gb. Even with existing low cost memory devices and high speed transmission networks, the amount of storage space and processing power required by the database to serve multiple users accessing the network to store and retrieve data is not scalable. Therefore a mechanism to embed multiple messages of differing security requirements into a one-dimensional signal which is also robust to storing in a compressed format is implemented.

The following procedure (shown in figure 4.1) has been tested in the experiments:

1 . An EEG signal of 100s, **c** is transformed using DWT/ICA, $\mathcal{T}$.

2 . WMs 1 to 4 are embedded in the selected transform domain components of **c**, $\mathcal{C}$.

3 . The signal with the embedded data $\tilde{\mathbf{c}}$ is reconstructed by applying the inverse of the applied transform, $\mathcal{T}^{-1}$ to the watermarked $\mathcal{C}$, $\tilde{\mathcal{C}}$.

4 . As a compression attack, $\tilde{\mathbf{c}}$ is modelled as an autoregressive process of order six. The compression method based on the AR process has been discussed in chapter 3.

5 . The attacked watermarked EEG signal, $\hat{\mathbf{c}}$ is reconstructed at the receiver using the prediction co-efficients polynomial A, the N initial co-efficients of $\tilde{\mathbf{c}}$ and the error co-efficients e_p.

6 . Recovery of the embedded message for both the DWT and ICA methods is attempted. The bit error rate *dist* is calculated for each of the embedded messages to find the corresponding bit error rate.

7 . Steps 2 to 6 are repeated a hundred times and the mean bit error rate is calculated for each message. The mean bit error rate is used to verify the suitability of each

technique in an eHealth system.

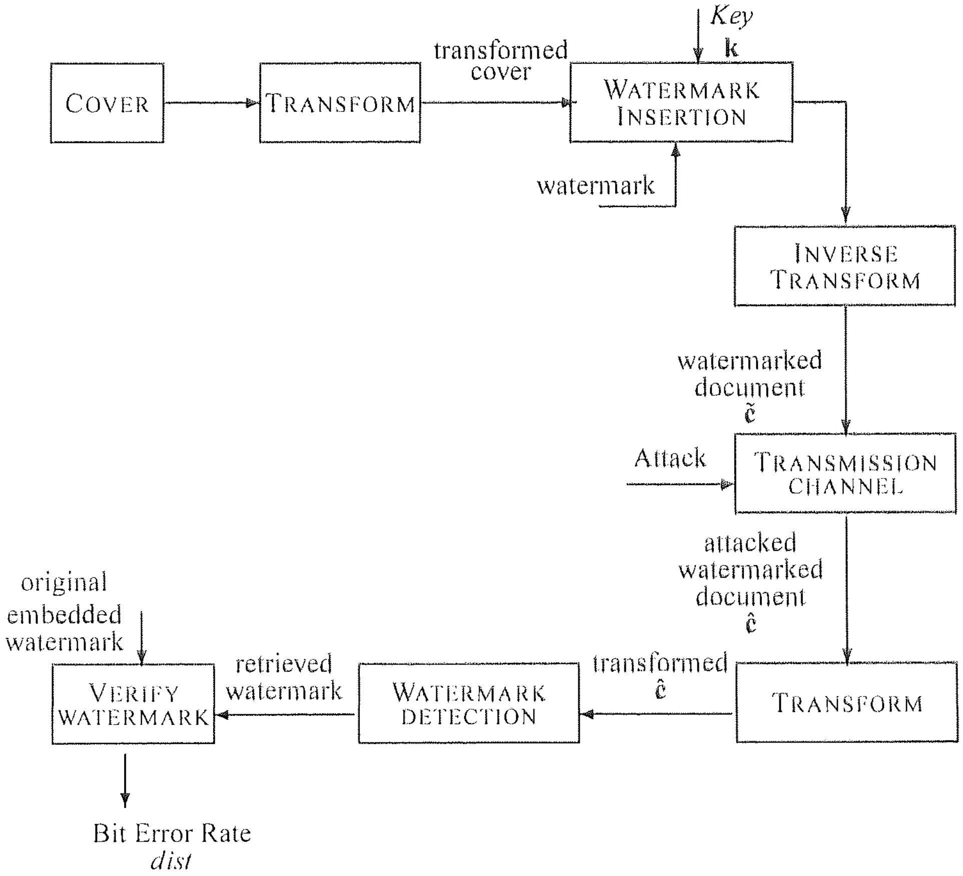


Figure 4.1: Block diagram of watermarking system.

4.4 Deriving Multiple Channels to Embed Multiple Watermarks

In order to embed the four watermarks of differing security requirements, four channels with matching characteristics are to be derived from the one-dimensional EEG. It will be shown how the two methods DWT and ICA provide different channels of different properties.

4.4.1 DWT based method of deriving multiple channels

In chapter 2 a discussion of how multiple level wavelet decompositions of a one-dimensional **c** are derived is provided. The wavelet transform of a one-dimensional signal generates two wavelets d_a representing the scale co-efficients of **c** and d_b the translation co-efficients of **c**. The two wavelets are derived using a suitable mother wavelet ψ. ψ here denotes the Haar wavelet. A detailed description of the wavelet decomposition of **c** using the Haar wavelet has been discussed in chapters 2 and 3. $C_{a_{i+1}}$ and $C_{b_{i+1}}$ are obtained by applying the DWT transform to C_{a_i}.

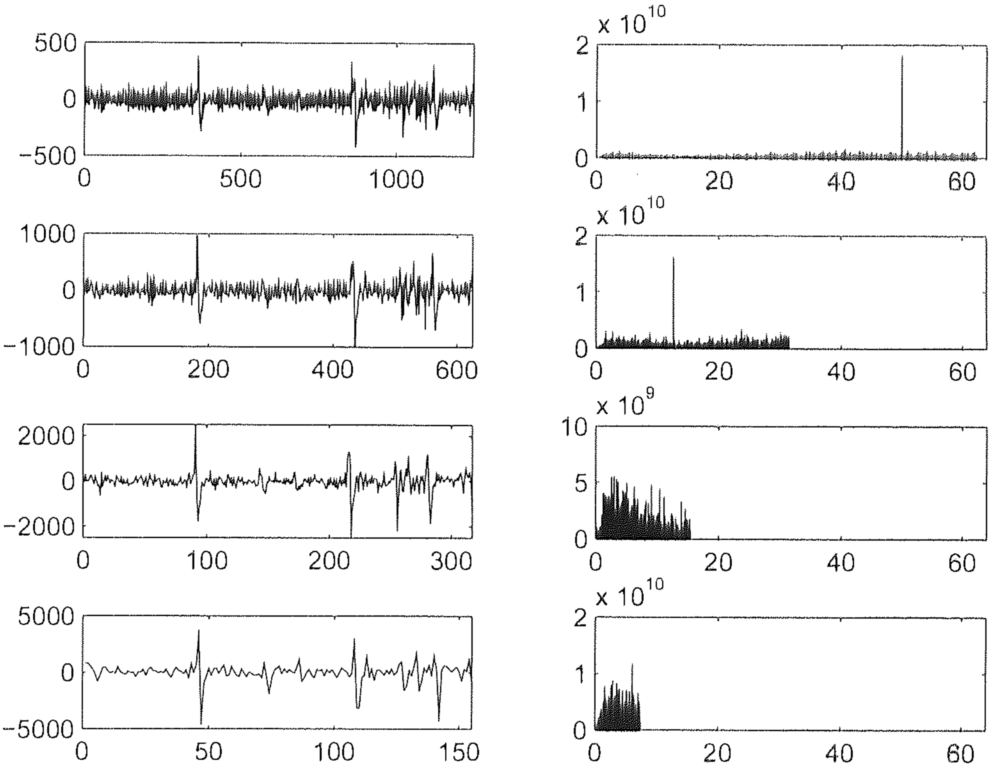


Figure 4.2: Four level wavelet decomposition of one-dimensional EEG signal and corresponding spectra.

C_a represent the slow moving components of the signal (the low frequency content) while C_b represent the high frequency components. C_{b_i} for each increasing level of decomposition represent signals of decreasing frequency.

Figure 4.2 depicts the detail coefficients obtained for each decomposition level one to four from top to bottom (left) with the corresponding frequency spectrum shown right. It is noticed the decreasing frequency content of C_b as the depth of decomposition increases.

It has to be noted that the EEG is being used as a cover for hiding sensitive patient related information. The transformation to the wavelets is done to obtain low frequency representations of the signal which are more robust to compression techniques and are preserved after compression. The application of DWT on **c** is not applied to derive the characteristics of **c**. Four levels of decomposition of **c** are obtained to embed the four watermarks.

Table 4.1 represents the energy content of the detail co-efficients at each decomposition level. The energy is calculated as

$$e_i = \frac{1}{Length(C_{b_i})} \sum_j C_{b_i}^2(j) \tag{4.2}$$

where i represents the decomposition level,

$$j \in \{1, Length(C_{b_i})\}. \tag{4.3}$$

Level	Energy
1	1.44e+04
2	7.04e+04
3	2.37e+05
4	1.00e+06

Table 4.1: Energy content of detail co-efficients.

From the table it is noted that C_b obtained for higher levels of decomposition have more energy compared to C_b obtained lower levels of decomposition. Since coefficients containing more energy are preserved after compression compared to the coefficients with lower energy content, the watermarks (semi-fragile watermarks) which need to be preserved after compression need to be embedded in the coefficients with higher energy content. $\mathbf{m}_4$ is embedded in the detail coefficients containing the maximum energy. Messages $\mathbf{m}_3$, $\mathbf{m}_2$, $\mathbf{m}_1$ are ordered in decreasing requirements of robustness and are therefore embedded in the detail coefficients obtained for decomposition levels 3, 2 and 1 respectively (with decreasing energy content).

4.4.2 ICA based method of deriving multiple channels

The one-dimensional EEG signal $\mathbf{c}$ used for the experiments in this chapter is taken from the 64 channel EEG recording used for the experiments in chapters 2 and 3. $\mathbf{c}$ in the ICA based watermarking method is transformed into an embedding data matrix $\mathbf{X}$ using delay embedding discussed in chapters 2 and 3. The minimum embedding window $\mathbf{EmbWin} =$ 83 sets the limit on the total number of sources p that can be estimated from the EEG.

$$[\mathrm{S}, \mathbf{W}] \overset{ICA}{\leftarrow} \mathbf{X}. \tag{4.4}$$

The rows of $\mathbf{S}$ represent the estimated independent sources and rows of $\mathbf{W}$ represent the independent components. Four sources $\mathbf{s}_{wm_{\mathrm{i}}}$, $\mathbf{s}_{wm_{\mathrm{ii}}}$, $\mathbf{s}_{wm_{\mathrm{iii}}}$ and $\mathbf{s}_{wm_{\mathrm{iv}}}$ are chosen from the l sources based on the spectrum of each source. A low frequency spectrum source contains a high information content of $\mathbf{c}$. The source with a high information content is more robust to compression compared to the source with a low information content. $\mathbf{s}_{wm_{\mathrm{i}}}$, $\mathbf{s}_{wm_{\mathrm{ii}}}$, $\mathbf{s}_{wm_{\mathrm{iii}}}$ and $\mathbf{s}_{wm_{\mathrm{iv}}}$ are ordered in terms of increasing information content represented as $\mathbf{s}_{wm_i}$ where i=[1, to 4]. For each watermark WM_i, $\mathbf{K}_i$ is the set of samples representing one of the four selected sources. $\mathbf{k}_i \in \mathbf{K}_i$ is randomly selected to embed the watermarks. The length of $\mathbf{k}_i$ is equal to WM_i.

It was shown how the delay embedding method is capable of extracting good estimates of the underlying sources $\mathbf{s}_i$ from a one-dimensional observation vector for biomedical signal analysis in chapters 2 and 3. Those four sources $\mathbf{s}_{wm_i}$ used to embed the four watermarks and their corresponding frequency spectra are shown in figure 4.3. The order and scale of the derived sources from the ICA do not bear any importance either on the rank of independence or frequency content of the estimated sources. The four watermarks require different levels of robustness to various forms of attacks. Differentiating between the various sources to estimate the capability of a source to survive an attack is therefore necessary. It was noted that the energy content of all the sources is normalised to one due to the whitening of the input to the ICA. Hence the energy content is not used as a distinguishing feature. The p sources were clustered into four different groups based on distances between the frequency spectra to distinguish the effect of embedding information in a particular source on the watermarked EEG.

The sources in each cluster have different spectra, hence the effect of a compression attack on a source in one cluster is different from the effect of compression on a source in a different cluster.

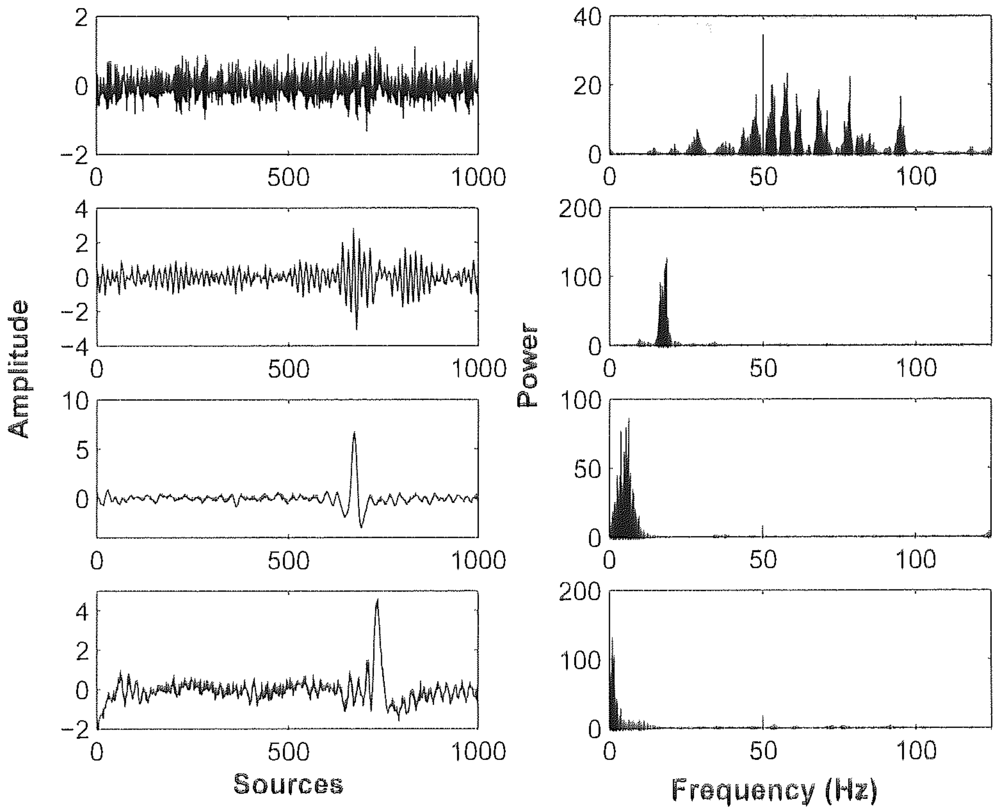


Figure 4.3: Estimated sources from one-dimensional EEG and their corresponding spectra.

Figure 4.4 is a dendrogram showing the four different groups of sources. The Euclidean distance between the spectra of sources was used as the separating factor. Since the EEG is a noisy data set, most of the derived sources have wide band spectra (representing noise) which is why cluster 1 has most of the sources. The sources in the other clusters are band limited as seen in figure 4.3. A representative of each cluster was used to embed each of the four watermarks.

4.5 Watermark Generation, Embedding, Transmission and Decoding

4.5.1 Watermark generation

Four different watermarks of varying levels of robustness requirement were generated as in [33]. The characteristics of each watermark in our case is different as we believe

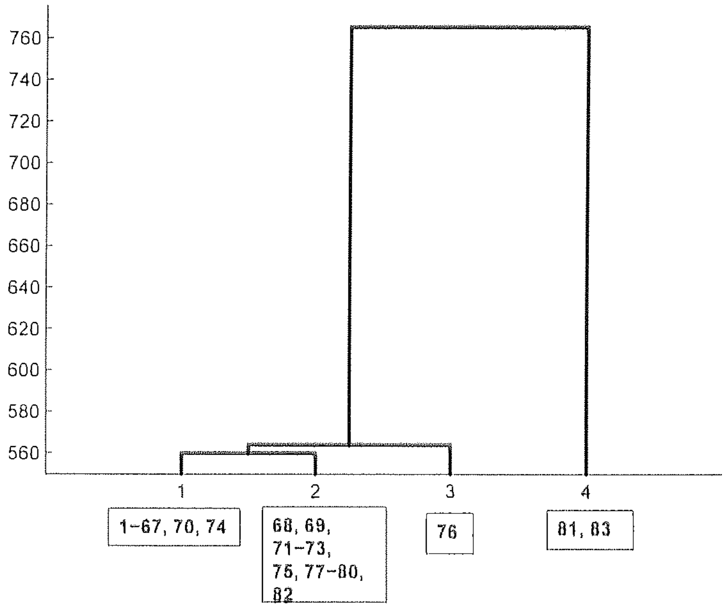


Figure 4.4: Dendrogram showing clustering of estimated sources.

that the personal details of the patient require the highest security. Any attempt by an intruder with the best possible resources and knowledge of the watermarking technique must not result in the personal details being recovered correctly. This was arrived at after discussion with various partners in Biopattern. A working document [29] which emphasises the importance of patient privacy was also produced after studying various legal cases that arised due to the implementation and use of medical databases. Details of the medical databases and the legal and ethical issues can be obtained from the reference within [29]. The four watermarks are ordered in increasing order of privacy and security requirement with WM_4 requiring the highest security.

As mentioned earlier, the embedded information distorts the cover work. The three characteristics of robustness, imperceptibility (capability of embedded information to remain indistinguishable from $\mathbf{c}$) and data rate (length of the watermark) are a trade-off against each other. Figure 4.5 shows the desired trade-off location for each watermark. The maximum length of each watermark is restricted to 301 bits. Due to the delay embedding of the one-dimensional EEG to construct the input matrix to the ICA (with a delay of one sample between two consecutive delay vectors), all the elements of an anti-diagonal are equal. In order to reconstruct the one-dimensional signal from the delay vectors only one element from each anti-diagonal is to be considered. Hence only one element on

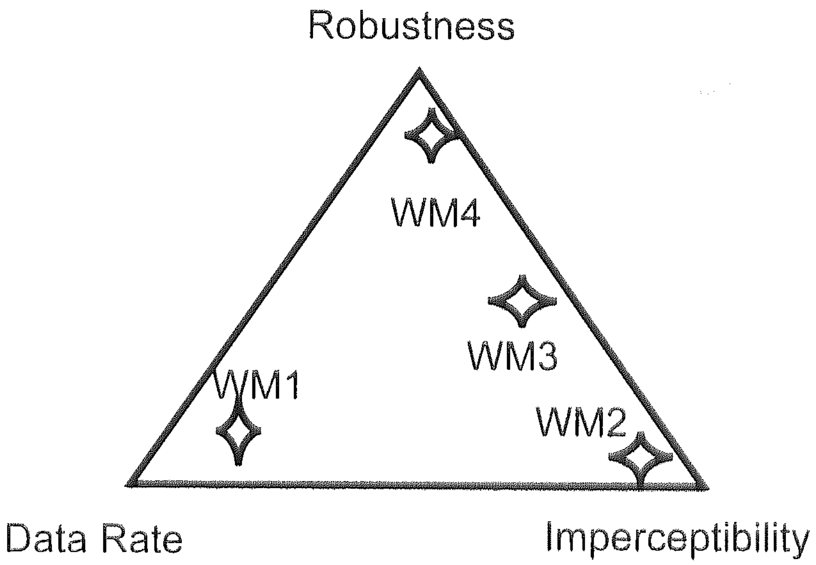


Figure 4.5: Desired characteristics of each watermark.

an anti-diagonal can be watermarked. This constrains the length of each watermark to a maximum of 301 values. Based on the position of each watermark in the trade-off triangle (defined by required robustness and imperceptibility) WM_1 has the maximum length and WM_4 has the minimum length. The length of the four watermarks is WM_1 = 300 bits, WM_2 = 150 bits, WM_3 = 75 bits, WM_4 = 37 bits. The length of WM_1 was limited by the ICA method. The estimated sources are of unit energy and equal length, but the energy level of the detail co-efficients obtained in the wavelets almost double with the depth of decomposition (Table 4.1) while the number of detail co-efficients is halved. The length of $WM_{2,3,4}$ was decided on the halving of the length of the wavelets.

4.5.2 Watermark embedding

As mentioned in chapters 2 and 3 the characteristics of the embedded watermark can be controlled in the 'QIM' method of embedding information by altering the value of δ. The watermark embedding function $\mathcal{F}$ is based on QIM which has been discussed in chapters 2 and 3. All the experiments are based on the fundamental concept that the information embedded in the low frequency components is capable of surviving compression attacks while it distorts the cover work $\mathbf{c}$ significantly. The number of quantisation levels were thus chosen to depict the robustness of the DWT/ ICA method under a compression attack.

DWT

In the DWT based watermarking method the watermarks embedded in C_{b_i} where i represents a higher level of decomposition are more robust to compression attacks compared to watermarks embedded in C_{b_i} where i represents a consequently lower level of decomposition. From table 4.1 it is noticeable that C_b of the fourth level decomposition have a higher energy compared to C_b of the other decomposition levels. Hence C_{b_4} is more suited to carry the message requiring the highest level of privacy and protection WM_4.

WM_i is therefore embedded in the translation co-efficients of C_{b_i}, where i=[1 to 4]. C_{b_i} represents the corresponding $\mathbf{K}_i$. $\mathbf{k}_i \in \mathbf{K}_i$ are chosen randomly to embed WM_i. The length of $\mathbf{k}_i$ is equal to the length of WM_i. The watermarked wavelet co-efficients $\tilde{C}_{b_i}$ are obtained as follows:

$$\mathcal{F}(WM_i, C_{b_i}, \mathbf{k}_i) \rightarrow \tilde{C}_{b_i}. \tag{4.5}$$

Applying the inverse of the DWT decomposition to the scale and translation co-efficients starting from the last level of decomposition to the first $\tilde{\mathbf{c}}$ is obtained.

$$DWT^{-1}(C_{a_i}, \tilde{C}_{b_i}) \rightarrow C_{a_{i-1}}. \tag{4.6}$$

where i takes values in the decreasing order starting from the highest decomposition level (for example in our experiments, decomposition level 4).

$$DWT^{-1}(C_{a_1}, \tilde{C}_{b_1}) \rightarrow \tilde{\mathbf{c}}. \tag{4.7}$$

ICA

The source containing signals of low frequency (see figure 4.3), was used to embed the WM_4 which requires the highest robustness to attack. The source containing higher frequency distribution was used to embed the fragile WM_1.

The embedding of the watermarks in each $\mathbf{s}_{wm_i}$ is conducted using QIM as in the DWT based method to obtain the watermarked source $\tilde{\mathbf{s}}_{wm_i}$.

$$\mathcal{F}(WM_i, \mathbf{s}_{wm_i}, \mathbf{k}_i) \rightarrow \tilde{\mathbf{s}}_{wm_i}. \tag{4.8}$$

The watermarked estimated sources matrix $\tilde{\mathbf{S}}$ contains the four watermarked sources $\tilde{\mathbf{s}}_{wm_i}$ and the $l-4$ unwatermarked sources. Applying $\mathbf{W}^{-1}$ to $\tilde{\mathbf{S}}$, $\tilde{\mathbf{X}}$ is obtained,

$$\mathbf{X} = \mathbf{W}^{-1} * \tilde{\mathbf{S}}. \tag{4.9}$$

The limitations of this method when used in a data hiding application have also been discussed in chapter 2. This is because data hiding applications require that $\tilde{\mathbf{c}} \sim \mathbf{c}$. The problem of the delay embedding method of constructing $\mathbf{X}$ when used in a data hiding application is revisited with a numerical example.

Numerical example:

$$\begin{aligned} \text{Input to ICA} &= \mathbf{X}, \\ \text{Separating matrix} &= \mathbf{W}, \\ \text{Mixing matrix} &= \mathbf{A}. \end{aligned} \tag{4.10}$$

One-dimensional cover $\mathbf{c}$ =[1 2 3 4 5 6];

$\mathbf{X}$ is constructed from $\mathbf{c}$ using the delay embedding technique with a delay of 1 sample.

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \end{bmatrix}. \tag{4.11}$$

The ICA method estimates the underlying sources $\mathbf{S}$ the separating matrix $\mathbf{W}$ and the mixing matrix $\mathbf{A}$ which is the inverse of $\mathbf{W}$.

$$\mathbf{S} = \begin{bmatrix} 0.4390 & 2.3645 & 4.2900 & 6.2155 & 8.1410 \\ 9.4939 & 24.1182 & 38.7425 & 53.3669 & 67.9912 \end{bmatrix}, \tag{4.12}$$

$$\mathbf{W} = \begin{bmatrix} 3.4120 & -1.4865 \\ 19.7548 & -5.1305 \end{bmatrix}, \tag{4.13}$$

$$\mathbf{A} = \begin{bmatrix} -0.4326 & 0.1253 \\ -1.6656 & 0.2877 \end{bmatrix}. \tag{4.14}$$

Let the watermarked sample be the 3rd sample in the first row and the watermarked sources $\tilde{\mathbf{S}}$

$$\tilde{\mathbf{S}} = \begin{bmatrix} 0.4390 & 2.3645 & 4.32 & 6.2155 & 8.1410 \\ 9.4939 & 24.1182 & 38.7425 & 53.3669 & 67.9912 \end{bmatrix}. \tag{4.15}$$

The watermarked input matrix is obtained as $\tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}}$

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1.0000 & 2.0000 & 2.9870 & 4.0000 & 5.0000 \\ 2.0000 & 3.0000 & 3.9500 & 5.0000 & 6.0000 \end{bmatrix}. \tag{4.16}$$

Only one sample across all the sources in a diagonal can be altered to carry a sample of the watermark since during reconstruction only one sample from each diagonal (top right to bottom left) of $\tilde{\mathbf{X}}$ is used to reconstruct the one-dimensional $\tilde{\mathbf{c}}$. Reconstructing the original signal from the overlapping segments therefore requires careful re-ordering (constructing $\tilde{\mathbf{c}}$ with the watermark modified samples) as mentioned in chapter 3. Hence method 2 of the reconstruction of the one-dimensional $\tilde{\mathbf{c}}$ from $\tilde{\mathbf{X}}$ discussed in chapter 2 is adopted.

4.5.3 Watermark transmission

The watermarked document $\tilde{\mathbf{c}}$ is corrupted during transmission due to intentional/common signal processing distortions η, resulting in an attacked watermarked document $\hat{\mathbf{c}}$.

$$\tilde{\mathbf{c}} + \eta \rightarrow \hat{\mathbf{c}}. \tag{4.17}$$

η in the experiments mentioned in this chapter refers to different signal processing attacks (compression, low pass filtering, addition of Gaussian noise and encoding using different number of bits). Figure 4.6 depicts the various parameters that are transmitted to the decoder in the two watermarking methods based on DWT and ICA.

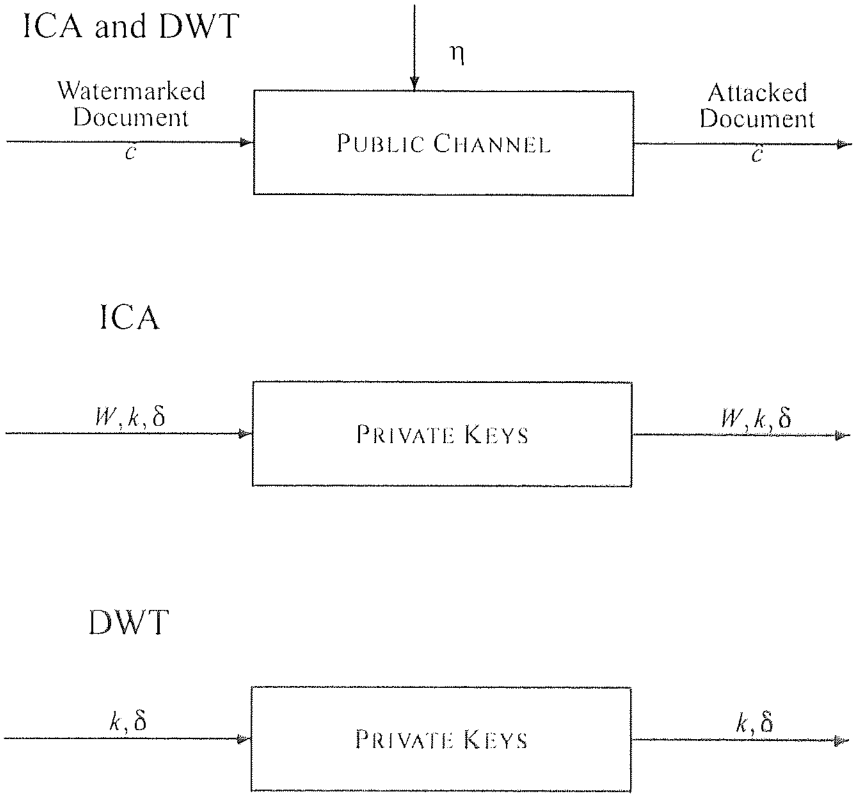


Figure 4.6: Transmission parameters for ICA and DWT watermarking systems.

The security of the embedded message in the DWT based method and the ICA based method has been discussed in chapter 3. The ICA based method requires the exact separating matrix $\mathbf{W}$ used by the encoder at the decoder to decompose the received $\hat{\mathbf{c}}$ into estimated sources [55].

It has been shown in chapter 3 how $\mathbf{W}$ acts as a key and prevents illegal estimation of the embedded information. This indicates that the level of security for the embedded watermarks is higher in the ICA compared to the DWT. Hence, though the data required by the decoder to reconstruct the embedded message (figure 4.6) in the case of the ICA based method is larger than the DWT, privacy and security to the embedded information is higher in the ICA approach.

4.5.4 Watermark decoding

$\hat{\mathbf{c}}$ is transformed using DWT or ICA to obtain the wavelet decomposition/estimated sources respectively. By applying the nearest integer level decoding technique (even for zero, odd for one) explained in chapter 1 an estimate of the embedded message is obtained.

The inverse of the watermark insertion and watermark generation process is applied sequentially to $\hat{\mathbf{c}}$ to obtain estimates of the embedded messages, $\hat{\mathbf{m}}_i$.

$$\mathcal{T}(\hat{\mathbf{c}}) \rightarrow \hat{C}. \tag{4.18}$$

$$\mathcal{F}^{-1}(\hat{C}(\mathbf{k})) \rightarrow \hat{WM}. \tag{4.19}$$

$$\hat{WM} \rightarrow \hat{\mathbf{m}}. \tag{4.20}$$

In the absence of the original unwatermarked $\mathbf{c}$ and the embedded messages $\mathbf{m}_i$ at the decoder, some prior knowledge about the type of $\mathbf{m}_i$ is required to assess the validity of the decoded message $\hat{\mathbf{m}}_i$. Since the embedded messages are mostly textual (denoting patient name, address etc.,) in the biomedical domain, retrieval of an intelligible hidden message is sufficient to provide authentication of the originality of $\hat{\mathbf{c}}$.

As discussed in chapter 2 the verification of the decoding process is conducted by calculating the Hamming distance, *dist* between the binary watermarks embedded in $\mathbf{c}$ and the estimated watermarks from $\hat{\mathbf{c}}$.

$$dist = \sum(WM \oplus \hat{WM}). \tag{4.21}$$

The larger the difference (value of *dist*), the greater the attack.

The transmission parameters the watermarked EEG signal, the coefficients of **W** were tested for various attacks. The attacks on the watermarked EEG signal constituted compression, low pass filtering, quantisation using differing number of bins and encoding using different bit rates and, addition of Gaussian noise. The attack on the parameter **W** constituted and estimation of the values of **W** by applying the ICA to different input matrices. The input matrices were constructed from different EEG signals using the delay embedding method (the construction of the input matrix has been discussed in chapter 1).

Figure 4.7 shows the result of using different **W** at the watermark detector to retrieve the embedded watermark. It can be seen that the watermark retrieved using the **W** used at the watermark embedder goes to zero while the bit error rate for the retrieved watermark using the estimated **W** remains around 50%. Hence we conclude that the true **W** is required to retrieve the embedded message and therefore forms a secret key.

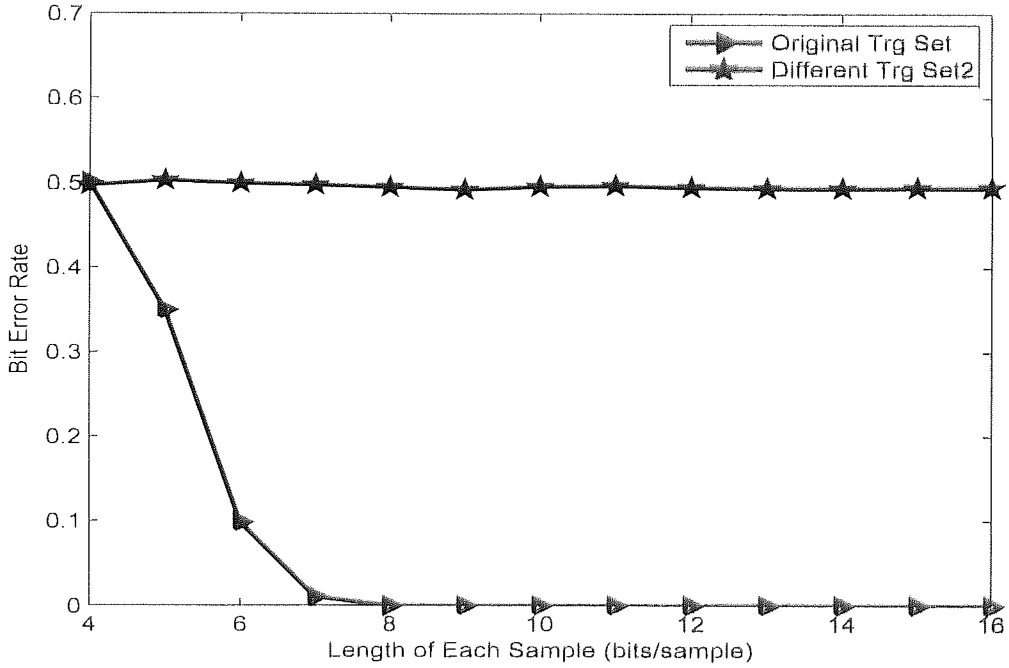


Figure 4.7: Bit error rate for watermarks retrieved by using different separating matrices.

4.6 Results

Figure 4.8 denotes the distortion due to the four watermarks embedded using the DWT based approach and the ICA based approach. It is a plot of the **c**, the recorded $\mathcal{D}_{Emb}$ due to the embedding of the four watermarks WM_i where i=[1 to 4] in $\mathbf{c}$ using the DWT approach followed by the $\mathcal{D}_{Emb}$ due to the ICA method respectively. $\mathcal{D}_{Emb}$ is relatively insignificant compared to the dynamic range of $\mathbf{c}$. It shows that the embedded watermarks should not influence the decision of the clinician viewing $\tilde{\mathbf{c}}$. The signal ($\mathbf{c}$) to noise ($\mathcal{D}_{Emb}$) ratio of $\tilde{\mathbf{c}}$ using the ICA technique was 42.86dB and in the DWT method 48.79dB. This SNR was calculated for the largest value of δ required to decode the watermarks with zero error. The DWT method has a lower value compared to the ICA but the distortion due to both the approaches is less than 0.05% of the dynamic range of the unwatermarked signal. This is due to the ratio of number of samples available to carry the watermark to the number of bits of the watermark. The ratio in the case of DWT is higher compared to the ICA. The performance of both the ICA and DWT in terms of imperceptibility can be regarded to be acceptable.

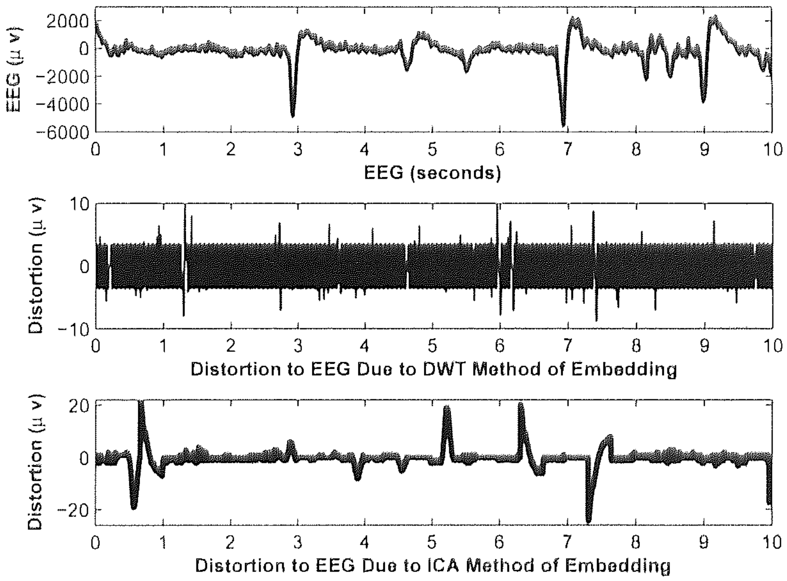


Figure 4.8: Unwatermarked EEG and distortion due to embedding multiple messages.

The results obtained for the various attacks on watermarked signal are shown in figures 4.9, 4.11, 4.13, 4.15 for the watermarks embedded using the DWT based method for different attacks such as compression, encoding with different bit rates, low pass filtering and addition of noise respectively. Similarly figures 4.10, 4.12, 4.14, 4.16 for the watermarks embedded using the ICA based method for different attacks such as compression, encoding with different bit rates, low pass filtering and addition of noise respectively.

The value of $\mathcal{R}$ in both the methods of embedding information remaining equal totalling 5.62 bits per second of the EEG, the number of quantisation levels for both the DWT and the ICA are maintained the same. The watermark reconstruction error is calculated by comparing the retrieved watermarks at the decoder to the original embedded watermarks. The Hamming distance between the watermark retrieved and the watermark embedded represents the bit error rate. From figures 4.9, 4.10 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 the bit error rates in the DWT based approach show that the watermarks are recovered in an incorrect order differing widely from the assumptions of robustness and security used during embedding the information. In the case of the ICA based method all the four watermarks are recovered in the correct order or as per the assumptions of robustness used at the watermark embedder.

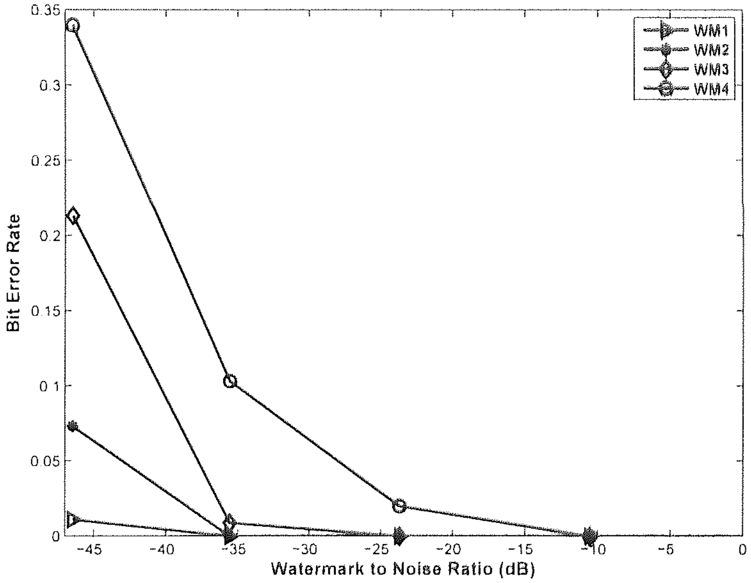


Figure 4.9: Decoded error for each watermark using DWT for compression attack. It should be noted that the robustness of the watemark four against an attack is larger than that of the fragile watermark one but watermark one is robust compared to watermark two and three. This shows that watermark one which should be the least robust among all the four watermarks is more robust compared to watermark two and three.

The experiments were conducted for 22 EEG signals taken from different data sets (EEG from patients undergoing seizures and EEG with no abnormal activity). Since the EEG contains multiple sources, in the case of EEG recordings with no abnormality we observed that there were no correlation between the signals recorded at different nodes. In case of an EEG recording with epileptic events there was a huge correlation between the

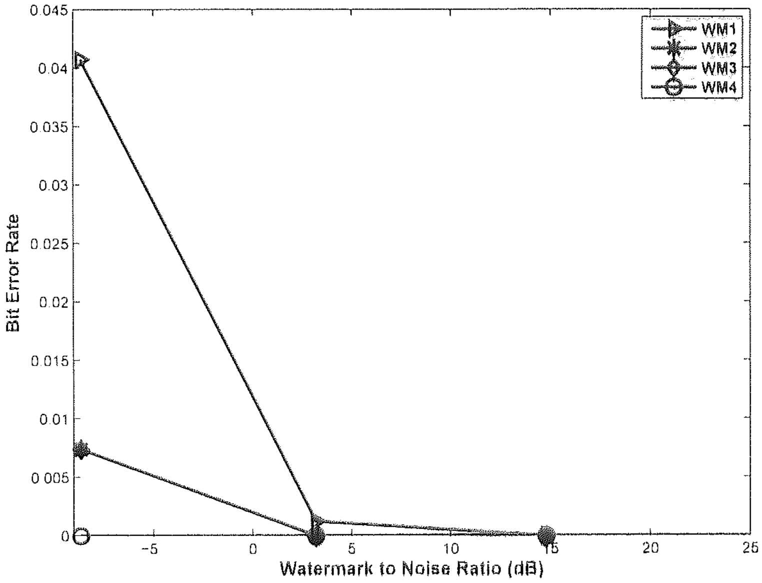


Figure 4.10: Decoded error for each watermark using ICA for compression attack. It should be noted that the robustness of the watemark four against an attack is larger than that of the all the other three watermarks and watermark one is the most fragile.

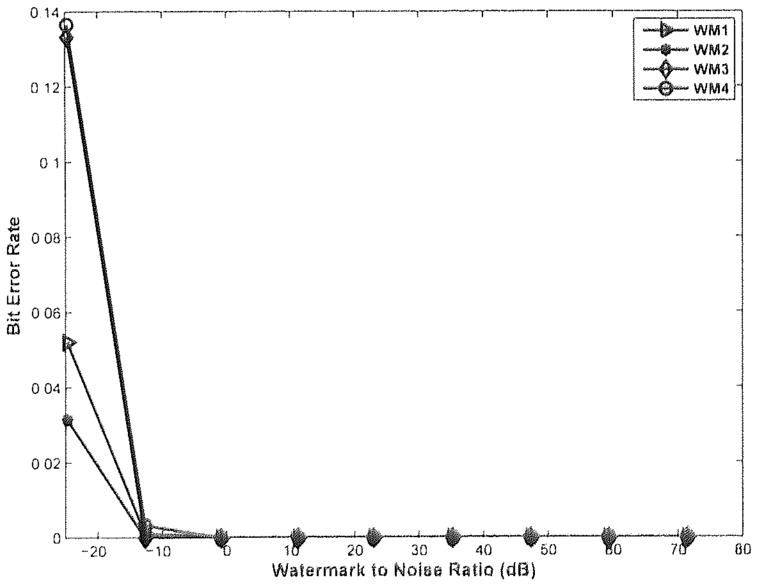


Figure 4.11: Decoded error for each watermark using DWT for requantisation attack. It should be noted that the four watermarks are not retrieved according to the design(in the order of robustness).

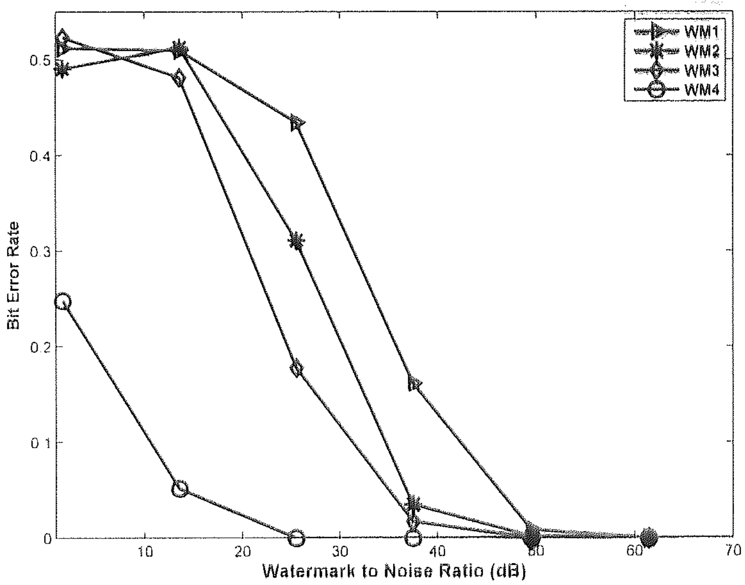


Figure 4.12: Decoded error for each watermark using ICA for requantisation attack. The watermark four has more robustness compared with the other watermarks. The four watermarks are retrieved in increasing order of robustness as designed.

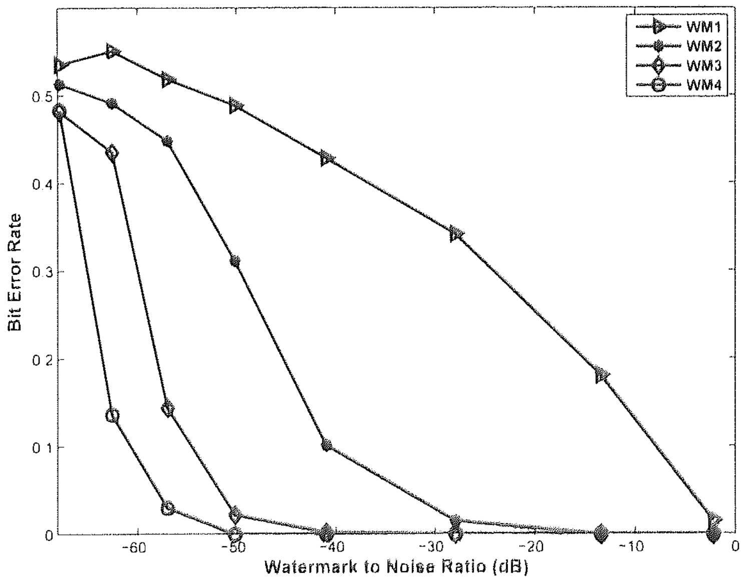


Figure 4.13: Decoded error for each watermark using DWT low pass filtering attack. All the four watermarks are retrieved in the correct order of robustness.

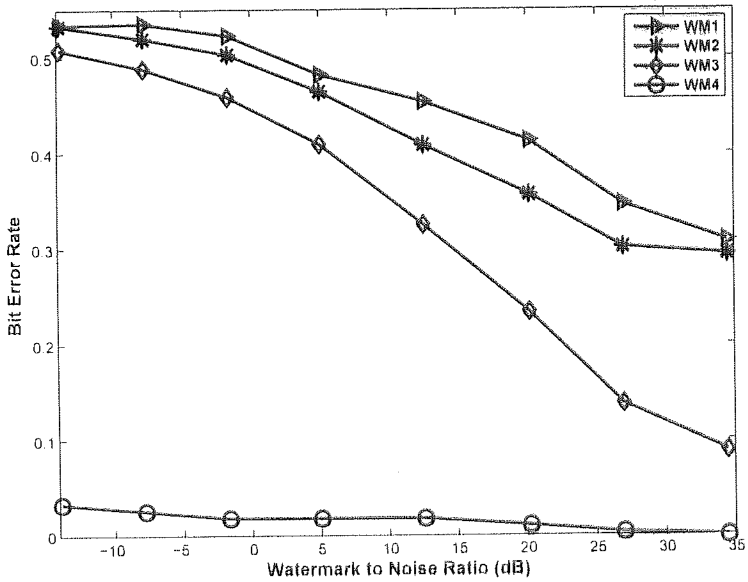


Figure 4.14: Decoded error for each watermark using ICA for low pass filtering attack. It should be noted that the robustness of the watemark four against an attack is larger than that of the all the other three watermarks and watermark one is the most fragile.

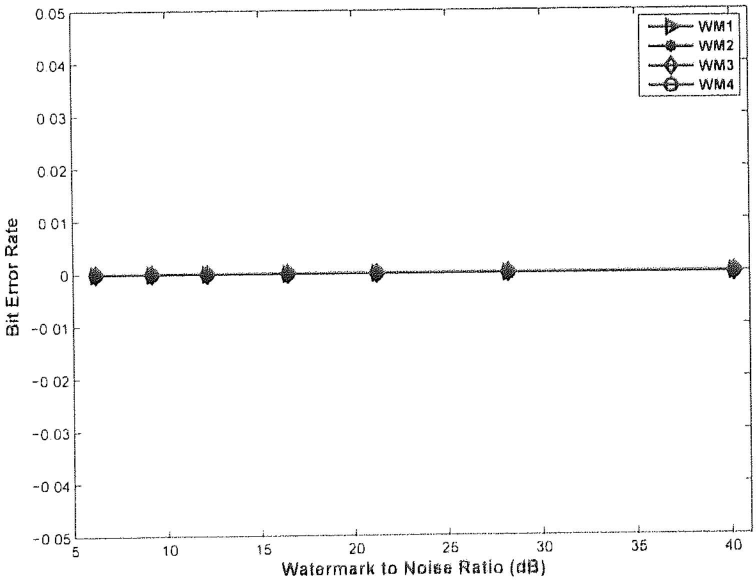


Figure 4.15: Decoded error for each watermark using DWT for additive noise. The watermarks embedded are more robust to additive noise compared with the ICA based approach.

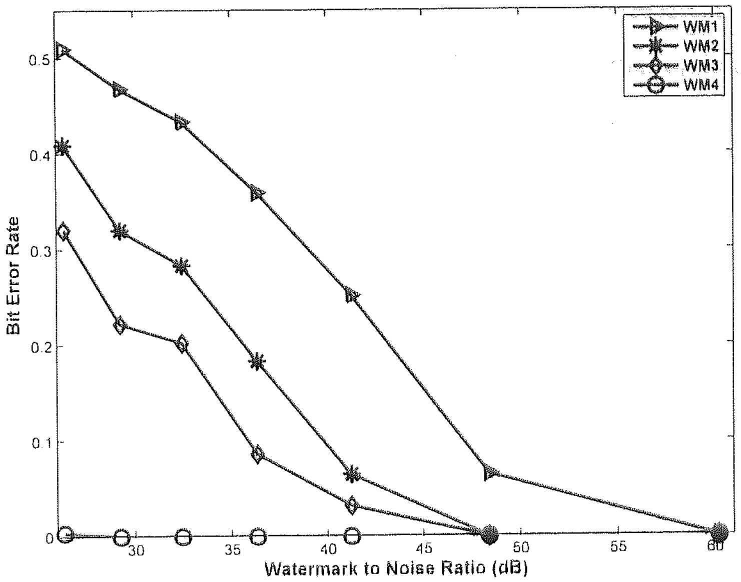


Figure 4.16: Decoded error for each watermark using ICA for additive noise. It should be noted that the robustness of the watemark four against an attack is larger than that of the all the other three watermarks and watermark one is the most fragile.

signals recorded at different nodes. Since in our experiments the recording taken at each node (one-dimensional EEG) is considered as an individual cover text, the correlations or abnormalities did not affect the actual watermark embedding/decoding method, or results. Variations in the bit error rates were observed for different one-dimensional EEG signals but the difference was less than 10%. It should be noted that by altering the quantisation index, this difference could be altered.

Designing embedding methods suitable for every type of attack is not possible. And embedding techniques robust to severe attacks is not required for medical applications. If the attack is severe **c** (such as the EEG in our example) will be compromised which renders it useless for diagnosis. As privacy of personal data in the medical record is the main concern in rolling out the EPHR, an approach which requires more information to retrieve the embedded information (see result of figure 4.7) is better suited. As shown from our experiments, the ICA approach provides more control over security and distortion for a given data rate.

4.7 Conclusion

From the experiments and the results obtained it can be seen that embedding of multiple watermarks of differing characteristics into a one dimensional cover is possible using both the DWT and the ICA based approach. In terms of the trade-off obtained between the two methods of DWT and ICA, the ICA performs slightly better, privacy of the sensitive personal information in the EPHR being the main concern, the ICA is advantageous. The ICA based method is better in providing security and privacy to personal data in a medical record due to the amount of information required to enable correct decoding of the embedded information. The transform co-efficients being fixed in the DWT based method an attacker can erase the embedded information if the watermark embedding method is known. This is not possible in the case of the ICA based method. An attacker cannot erase the embedded information even if she has knowledge about the exact watermarking method. The original unwatermarked cover is required to obtain the exact transform coefficients. This is the advantage of the ICA method over the DWT. The work presented in this chapter was part of the proceedings of a conference on artifical neural networks [59].

The next chapter considers embedding multiple watermarks of similar characteristics as opposed to the embedding of dissimilar characteristics watermarks. This is required if multiple watermarks are used as a log of access of the medical record. The information provided by each watermark is the same. The problem in such watermarking applications lies in deriving the order of the embedded watermarks. This problem of ordering of the multiple embedded watermarks at the decoder is considered in the next chapter.

5 SEQUENCING MULTIPLE WATERMARKS

CONTENTS

Embedding messages such as the personal details of a patient into biomedical data to protect the privacy of the individual has been presented in the previous chapters. It was also shown that the ICA performs better compared to the DWT in terms of security. The multiple watermarks $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4$ embedded in chapter 4 had differing requirements of robustness and security. In this chapter it will be shown how similar characteristic multiple watermarks $\mathbf{m}_{sim_1}, \mathbf{m}_{sim_2}, \mathbf{m}_{sim_3}, \mathbf{m}_{sim_4}$ being embedded in an EPHR could provide a tracing mechanism. One benefit of embedding multiple watermarks of similar robustness that this chapter will demonstrate, would be to serve as a temporal log of recent activity on an EPHR. A log of the clinicians who have accessed the medical record is an important requirement in an EPHR. It promotes confidence among the members of the public that their confidential records are not being accessed by unauthorised personnel. This chapter reveals how such an access mechanism log can arise from our data hiding algorithm which is currently not possible by most other data hiding techniques. To achieve this requires the order of sequence in which each watermark is embedded, to infer whose watermark was embedded last, for example. This property is hard in the hierarchy of data hiding algorithm.

To determine the sequence in which multiple watermarks have been embedded requires an asymmetry in the temporal embedding process. The one-dimensional EEG signal is decomposed into a plurality of components and the different watermarks are embedded at different instants of time. It is necessary to determine which signal components are likely to be substantially unaffected by the expected degradations due to the various signal processing applications during the embedding of each of the watermark. It is also necessary to identify the signal components which are independent of each other such that altering one of the signal components does not affect the other signal components. In the absence of asymmetry (when different signal components are not statistically independent), obtaining the sequence is not possible as embedding a second message into a watermarked cover could destroy the first message. In the case of all the watermarks being largely recovered correctly a mechanism to establish the order of the watermark embedding is required.

In the event of embedded messages (details of personnel who have accessed the record) being used as a log of access, an unauthorised person who has accessed the record will attempt to delete her details. This corruption of the embedded message is possible by overwriting the information by embedding more information or destroying $\tilde{\mathbf{c}}$ to a large extent as already noted in the chapter 3. The problem with watermarked systems arises in the presence of a capable, determined attacker equipped with the necessary tools and knowledge required to detect the presence of, or destroy the embedded information. This attacker may want to erase any trace of unauthorised access. But as seen in chapter 3, the ICA method with its sensitivity to input data creates a barrier to unauthorised deletion of the embedded watermark without major distortion of $\tilde{\mathbf{c}}$.

This chapter investigates whether multiple watermarks can be embedded into a cover work at different instants of time, and recovered in sequence. The ICA method of estimating independent sources described in chapter 2 will be utilised to embed multiple watermarks (one watermark in one source). It will be shown that by selecting the sources based on certain criteria the multiple embedded messages can be retrieved and the order of embedding established. This ability is acknowledged to be extremely difficult ([15] and references within), and no other equivalent successful method for watermark sequencing exists in the literature.

5.1 Embedding Multiple Watermarks

As noted in the previous chapters a watermarking system is designed by an appropriate choice of domain (time, frequency) of $\mathbf{c}$, $\mathcal{T}$ used and $\mathcal{F}$ for a given range of η (compression, filtering, scaling, cropping, rotation, additive noise) based on the application and use of $\mathbf{m}$. All the watermarks used as a log of access are equally important as they represent the record of people who have accessed the EPHR. Hence the watermarking system should be ideally capable of providing multiple channels (to carry multiple messages) bearing the same properties (robustness and capacity). The data hiding method derived in this thesis based on the independent components is the most suitable of all transform domain based data hiding techniques since it derives multiple groups of sources for a given set of observations (see chapter 2) with specific useful properties. The derived sources can be easily clustered to obtain a set of channels to carry multiple watermarks of equal robustness and capacity. The data used as an examplar is a single channel EEG signal which is transformed into independent sources to obtain multiple potential embedding components. Cox et al [24] argue that binary watermarks are comparatively easy to modify compared to watermarks derived from $\mathcal{N}(0,1)$ under an attack. For example if $\mathcal{F}$ is based on QIM, distorting $\tilde{C}(\mathbf{k})_j$ by $\delta/2$ will reverse the embedded bit value. This property of binary watermarks embedded using QIM based techniques is utilised to derive the fragility of a message when it is attacked by the addition of more messages into the cover.

5.2 Embedding Information in a Single Channel EEG Using ICA

The application of the independent component approach to a single channel EEG signal has been discussed in chapter 2. The EEG used for the experiments in this chapter is taken from an EEG recording of 64 channels. The EEG is sampled at a frequency of 250Hz and each sample represented by a 16bit unsigned integer. The delay embedding method of constructing the input matrix $\mathbf{X}$ from a one-dimensional $\mathbf{c}$ described in chapter 2 is implemented. Given a delay of one sample between two successive vectors, the size of the embedding window is calculated at the ratio of the sampling frequency to the frequency of the slowest signal component. This results in $\mathbf{EmbWin} = 83$. This embedding window defines the upper bound on the number of sources $\mathbf{s}$. Embedding multiple watermarks of varied characteristics $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4$ in different souces $\mathbf{s}_{wm_i}$, $\mathbf{s}_{wm_{ii}}$, $\mathbf{s}_{wm_{iii}}$ and $\mathbf{s}_{wm_{iv}}$ and the requirements for the choice of a source $\mathbf{s}_{wm_i}$ was discussed in chapter 4.

Figure 5.1 shows a selection of six sources from the 83 sources with different spectra. The remaining sources are spectrally equivalent replicas of one of the six sources shown in figure 5.1. The value of $\mathbf{EmbWin}$ is based on the knowledge of the characteristics of the probable underlying sources mixed linearly to obtain $\mathbf{c}$. $\mathbf{EmbWin}$ is not defined by a knowledge of the actual number of underlying sources. The delay embedding method hence results in estimates of clusters of sources with distinct spectra. In particular, figure 5.1 shows that all the sources obtained are not identical time-delayed versions of each other. Compression attacks based on filtering usually filter out the noise to obtain a better representation of the data under observation. Hence embedding information in the low and middle frequencies of a signal preserves the embedded message after compression. But embedding information in the low frequency regions of $\mathbf{c}$ increases the value of $\mathcal{D}_{Emb}$. This is because the low frequency components of the signal contain most of the information of the signal and hence have a high power content. Instead embedding information in the middle frequency representations of $\mathbf{c}$ reduces the value of $\mathcal{D}_{Emb}$ and also survives compression attacks. Since each of the sources has a distinct spectra, each of the source provides a different level of robustness against different attacks and distortion.

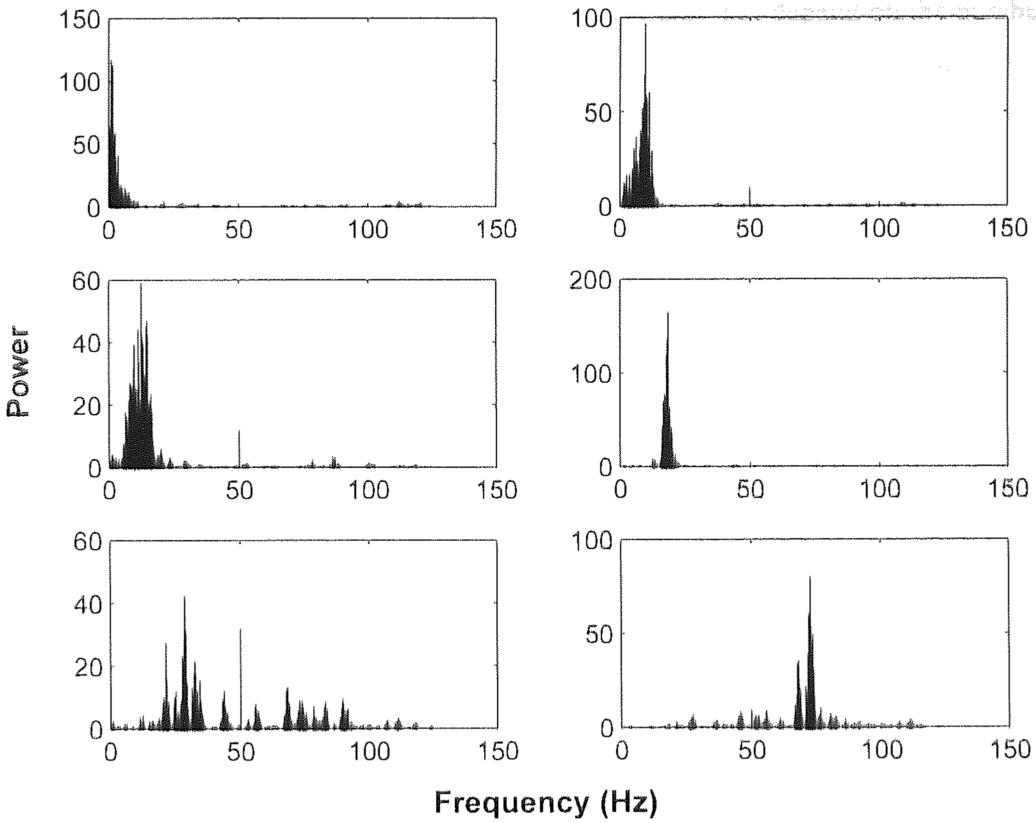


Figure 5.1: Six sources with different spectra.

5.3 Source Clustering

Extraction of independent sources $\mathbf{s}$ from a time series $\mathbf{c}$ by the embedding method is a redundant problem in that multiple solutions exist. This leads to the extraction of trivially distinct sources (such as differing in phase only for example). Therefore, clustering of functionally equivalent sources will need to be considered. The characteristics of the source determine the fidelity and robustness of the embedded watermark in that source. Therefore obtaining appropriate clustering is important when embedding multiple watermarks in a single channel signal.

Since the single channel signal does not provide information about the probable number of sources $\mathbf{s}$ and their characteristics to the watermark encoder or decoder, finding the number of clusters, N_{clust} is a blind operation. When the number of watermarks are known and are different in characteristics, the number of clusters can be decided based on the number of watermarks as in chapter 4. But 'if' the characteristics of all the watermarks are the same, then the choice of the number of clusters does not depend on the number of the watermarks. This is because the sources in each cluster will have similar spectra or power, but not exactly the same. These minor differences in the spectra/power of the different sources in a cluster could lead to differing levels of robustness of the embedded watermarks.

5.3.1 Similarity of time series data

Time series data are a necessary form of data in a multitude of applications for example weather forecasting and financial forecasting. Deriving the similarity between time series hence is important for many applications and not just biomedical data. Obtaining exact matching between two time series is not a practical proposition. Differences between time series could arise from measurement defects to differences in scale and time shifts, to differences in generative nosie sequences.

Distance metrics are defined to be non-negative, symmetric and obey the triangle inequality [34]. Identifying and clustering of sources using their time-domain representation is non-trivial and time-consuming due to the unspecified length of each source (being a time series) and non-availability of a proper distance measure. Some of the sources obtained from the ICA, though similar in terms of their frequency content or information content, may be scaled differently, rotated or shifted in time and so are trivially dissimilar.

Let us define a simple distance measure, $\mathrm{d}(\mathbf{s}_a, \mathbf{s}_b)$ given by the average of the difference between two time series $\mathbf{s}_a$ and $\mathbf{s}_b$, of equivalent length, $\mathrm{N_{ts}}$:

$$\mathrm{d} = \frac{\sum(|\mathbf{s}_a - \mathbf{s}_b|)}{\mathrm{N_{ts}}}. \tag{5.1}$$

$\mathrm{d}(\mathbf{s}_a, \mathbf{s}_b)$ is positive $\mathrm{d}(\mathbf{s}_a, \mathbf{s}_b) \geq 0$ and symmetric; $\mathrm{d}(\mathbf{s}_a, \mathbf{s}_b) \equiv \mathrm{d}(\mathbf{s}_a, \mathbf{s}_b)$ From Parseval's theorem, the energy of a signal in the time domain is equal to the energy in the frequency domain. The mean energy of a signal in the time domain is equal to the mean energy of the signal in the frequency domain. Let $\mathbb{E}$ represent the mean of a signal. Hence the distance between two time series in the time domain is the same as their distance in the frequency domain. Let $f_{s_a}(w)$, $f_{s_b}(w)$ represent the spectrum of the time series $\mathbf{s}_a$, $\mathbf{s}_b$.

$$\mathrm{d}(\mathbf{s}_a, \mathbf{s}_b) = \mathbb{E}[|\mathbf{s}_a - \mathbf{s}_b|] = \mathbb{E}[|\mathbf{s}_b - \mathbf{s}_a|] = \mathrm{d}(f_{s_a}(w), f_{s_b}(w)). \tag{5.2}$$

[83] used the Fourier transform domain representation of the time series to derive the similarity between two time series data. Popivanov et al [81] apply the wavelet transform to the time series and select a subset of the wavelet coefficients which represent the feature space of the time series to estimate the similarity. These methods are mainly utilised in data mining applications. Based on Parseval's theorem the similarity between the frequency spectra of the sources is used as an identifying mechanism to separate the sources into clusters.

Visual inspection and classification of the estimated sources is the most commonly used technique when the ICA is used as a signal analysis tool. But it is a tedious and uneconomical method. Various clustering algorithms exist which group the estimated sources into distinct clusters using different distance measures. One of the commonly known and used metrics is the Euclidean distance measure. The main disadvantage of clustering methods using Euclidean distance is their inability to classify delayed versions of the same signal as being similar. Ordinal analysis methods based on rank order and structure order of the time series override the problems in the classification of a set of signals containing subsets of time and scale shifted versions of the same signal.

Ordinal analysis of time series data has been studied by [8] to define the structure, obtain characteristic frequencies, and identify the time-dependence of future sample values of a time series. The rank order of a sample of a time series $\mathbf{s}$ of length $\mathbf{N}_{ts}$ at a given time t, $\mathbf{s}(t)$ is the number of samples of $\mathbf{s}$ that are lesser in value to $\mathbf{s}(t)$. The ordinal series derived from a time series hence provides an alternate approach in determining the relationship of different sample values obtained at different instants of time. The distance measures based on ordinal series will provide a different and probably better clustering of the estimated sources compared with the Euclidean distance measures, where the samples of two time series at the same time instant are compared.

In this chapter the given set of estimated sources are grouped using hierarchical clustering based on the Euclidean distance measure and grouped using ordinal methods. A comparison of the clustering due to the two methods is presented in the following sections.

5.3.2 Clustering using hierarchical classifier

The hierarchical classifier derives clusters based on the spectral distance measured between each pair of sources. The hierarchical clustering algorithm is used as it finds successive clusters based on previously established clusters. The hierarchical clustering results in a tree structure with each leaf representing an individual source. It does not require the number of clusters to be specified unlike partional clustering algorithms. This is advantageous as the underlying structure and the exact number of statistically independent sources (number of clusters) in an one-dimensional EEG are unknown. The clustering algorithm can be conditioned to cluster the sources based on whether the sources belonging to a cluster have a small distance between them or a large distance between them. The sources used to embed the watermarks in experiments implemented in this chapter are clustered such that the closer the sources are in the frequency domain the more probable they are in the same cluster.

Given a data matrix $\mathbf{S}_{m\times n}$, with row vectors $[\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3\ldots]$ the Euclidean distance d between two vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ is given as

$$\mathrm{d}_{ij}^2 = (\mathbf{s}_i - \mathbf{s}_j)(\mathbf{s}_i - \mathbf{s}_j)'. \tag{5.3}$$

5.3.3 Clustering based on ordinal analysis

In this subsection, a new method of clustering (designed by us), based on ordinal analysis of time series is described.

Ordinal analysis of time series data is based on the two properties of rank order and structure order. The definitions of rank order and structure order of a given time series $\mathbf{s}$ of length $\mathrm{N_{ts}}$ is given below [8].

Rank order: The rank $\mathbf{r}(i)$ of sample $\mathbf{s}(i)$ is given by the number of samples of $\mathbf{s}$ which are less than $\mathbf{s}(i)$ where i spans the length of $\mathbf{s}$.

$$\mathbf{r}(i) = \sum_{j=1}^{\mathrm{N_{ts}}} \Theta(\mathbf{s}(i) - \mathbf{s}(j)); \text{where } i \in \{1, \mathrm{N_{ts}}\} \& j \neq i \tag{5.4}$$

$$0 < \mathbf{r}(i) < \mathrm{N_{ts}} \forall i \tag{5.5}$$

$$\Theta(x) = \begin{cases} 1, & \text{if} x \geq 0 \\ 0, & \text{if} x < 0 \end{cases} \tag{5.6}$$

Structure Order matrix: The elements of the Structure Order matrix $\mathbf{SO}$ are the indicator functions comparing elements of $\mathbf{s}$.

$$\mathrm{h}_{ij} = \Theta(\mathbf{s}(i) - \mathbf{s}(j)). \tag{5.7}$$

The correlation coefficient between the rank order and structure order matrices of each time series $\mathbf{s}_i$ with every other time series $\mathbf{s}_j$, where i and j take values $\{1,83\}$ is calculated. The correlation coefficient is thresholded and the time series whose correlation coefficients are greater than the threshold are grouped together.

Rank order clustering:

Consider two time series $\mathbf{s}_a$ and $\mathbf{s}_b$ of equal length $\mathbf{N}_{\text{ts}}$. Let $\mathbf{r}_a$ and $\mathbf{r}_b$ represent the rank order of $\mathbf{s}_a$ and $\mathbf{s}_b$ respectively. Let $v(i)$ represent the sample whose co-ordinates are $(\mathbf{r}_a(i), \mathbf{r}_b(i))$. v denotes the spread of $\mathbf{r}_a(i)$ from $\mathbf{r}_b(i)$ on a square area defined by the $\mathbf{XY}$ axis.

$\mathbf{r}(i)$ of each $\mathbf{s}(i)$ is derived based on all $\mathbf{s}(j)$ where $j \neq i$. Therefore v is not affected due to a time-shift or scaling difference between $\mathbf{s}_a$ and $\mathbf{s}_b$. Let $x,y \in \mathbb{R}$ and c represent a constant. The correlation factor, $corr$ between $\mathbf{r}_a$ and $\mathbf{r}_b$ will be one when $\mathbf{r}_a$ and $\mathbf{r}_b$ are identical and v is linear as described in [8]. The correlation factor, $corr$ between $\mathbf{r}_a$ and $\mathbf{r}_b$ will be zero when $\mathbf{r}_a$ and $\mathbf{r}_b$ are orthogonal and v is uniformly spread across the square area.

$$corr = \begin{cases} 1, & \text{if } v = \{(\mathbf{r}_a(i), \mathbf{r}_b(i)) | x\mathbf{r}_a(i) + y\mathbf{r}_b(i) = c\} \\ 0, & \text{if } v = \{(\mathbf{r}_a(i), \mathbf{r}_b(i)) | x\mathbf{r}_a(i) + y\mathbf{r}_b(i) \neq c\} \end{cases} \tag{5.8}$$

Let $\mathbf{r}_i$; $i \in \{1,83\}$ represent the rank order of all 83 sources. Figure 5.2 (a) is the plot of $v(\mathbf{r}_1, \mathbf{r}_1)$. It is a straight line. It verifies that $\mathbf{s}_1$ is highly correlated with itself. Figure 5.2 (b), (c) and (d) are the plots of $v(\mathbf{r}_1, \mathbf{r}_2)$, $v(\mathbf{r}_1, \mathbf{r}_2)$, $v(\mathbf{r}_1, \mathbf{r}_3)$ respectively. The distribution of the values in the rank order in the subplots (b), (c) and (d), is uniform indicating that $\mathbf{s}_2$, $\mathbf{s}_3$, $\mathbf{s}_4$ are uncorrelated with $\mathbf{s}_1$. This absence of correlation was noticed for each pair of sources $\mathbf{s}_a$, $\mathbf{s}_b$ where $a \neq b$. Visual inspection of $\mathbf{s}_i$ $\forall i \in \{1,83\}$ showed that there exists groups of sources which are similar in structure. The sources in a group are time shifted versions of each other but this was not identified by correlating the rank order of the sources. Hence the frequency spectra of all $\mathbf{s}_i$ were considered instead of the time-structure.

Let $f_{s_i}(w)$ denote the spectrum of $\mathbf{s}_i$ and $\mathbf{r}_{f_i}$ the rank order of $f_{s_i}(w)$. Figure 5.3 is the comparison of the $\mathbf{r}_f$ of different pairs $f_{s_a}(w)$, $f_{s_b}(w)$ where $a \neq b$. Each subplot shows the comparison of the $\mathbf{r}_f$ of the two sources with the values of a and b, and the correlation $corr$ between $f_{s_a}(w)$, $f_{s_b}(w)$ mentioned above. $corr \in \{0,1\}$. The closer the value of $corr$

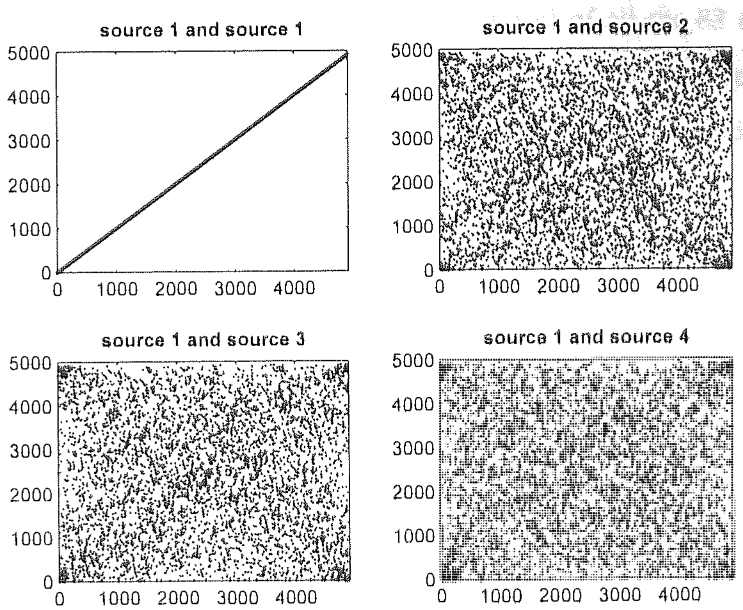


Figure 5.2: Rank order of source one compared with rank order of sources one to four in the time domain. It can be seen that the comparison of rank order of source one with itself is linear indicating that it is fully correlated. The comparison of the rank order of source one with rank order of sources two to four is spread in the square area whose axis represent the length of the sources. This shows that source one is independent of sources two, three and four.

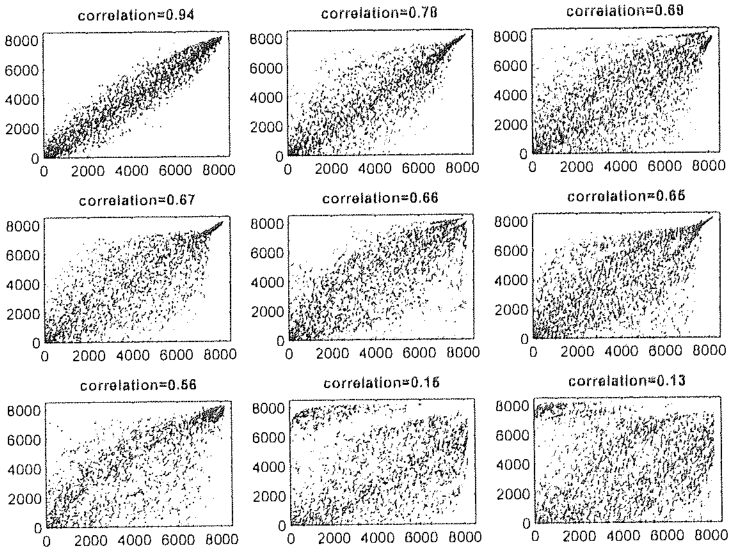


Figure 5.3: Comparison of the rank order of different pairs of spectra of the estimated sources. The correlation coefficient for each pair of spectra is shown above each plot. The plots have been arranged in decreasing order of correlation coefficients. It can be seen that the higher the correlation coefficient the closer the spread of the comparison to a linear plot. As the correlation coefficient decreases, the spread of the comparison of the rank order of the spectra, tends to become uniform in the square area whose sides are equal to the length of the time series.

to 1, the higher the dependence between the frequency spectra of $f_{s_a}(w)$ and $f_{s_b}(w)$ and the less the spread in the rank order matrix.

Figure 5.4 is the plot of the values of *corr* of $f_{s_i}(w)$ of all the 83 sources against each other using $\mathbf{r}_{f_i}$. A value of one indicates similarity (obtained for every $f_{s_i}(w)$ compared with itself) and a value of zero for two sources being uncorrelated in spectra and dissimilar. Since the values of *corr* are continuous values between zero and one, identifying clusters using this matrix is non trivial. Also certain applications require clusters of sources wherein the sources in a particular cluster are correlated to a percentage of the maximum correlation. In order to derive this clustering, the correlation matrix was reconfigured by thresholding the correlation index.

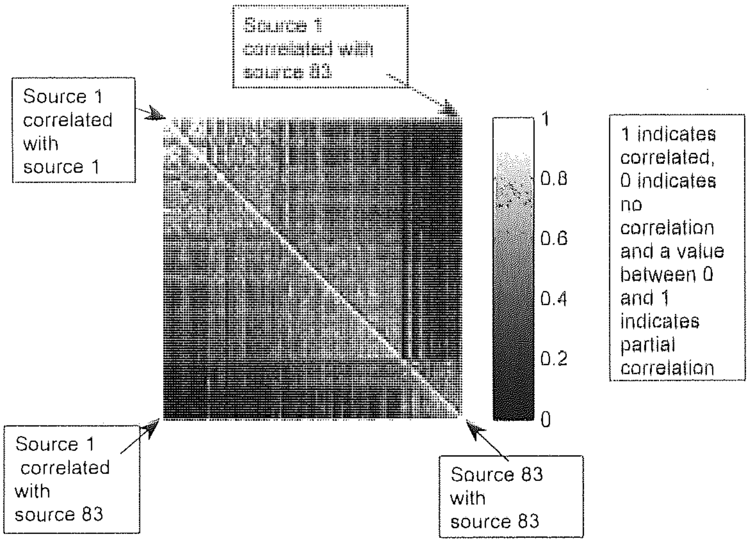


Figure 5.4: Correlation matrix obtained by comparing the spectrum of each source with the spectrum of all the other sources including itself. Each correlation coefficient represents a sample. A correlation coefficient of value one is shown as a white sqare and a correlation coefficient of value zero is shown as a black square. The correlation coefficient obtained for the comparison of the spectrum of each source with itself is a white square and represents the diagonal.

The subplots in figure 5.5 show the thresholded correlation matrices obtained for $\mathbf{r}_{f_i}$ of each of the estimated 83 $f_{s_i}(w)$ against $\mathbf{r}_{f_i}$ of every other $f_{s_i}(w)$ where $i \in \{1,83\}$. The thresholded value of *corr*, $corr_{\%}$ is printed above each plot. Pairs of $f_{s_a}(w)$, $f_{s_b}(w)$ where $a,b \in \{1,83\}$ which are correlated $\geq corr_{\%}$ mentioned are represented as one while the other pairs of $f_{s_a}(w)$, $f_{s_b}(w)$ are represented as zero. It can be seen that the higher the value of $corr_{\%}$, the less the number of pairs of $f_{s_a}(w)$, $f_{s_b}(w)$ sources which are similar to each other and vice versa. Using the plots in this figure helps find clusters of sources. For a low value of $corr_{\%}$ (≤ 0.6) the represention of *corr* has a large number of zeros and for $corr_{\%} = 0.8$, *corr* has a large number of ones. For $corr_{\%} = 0.7$ diagonal bars of zero

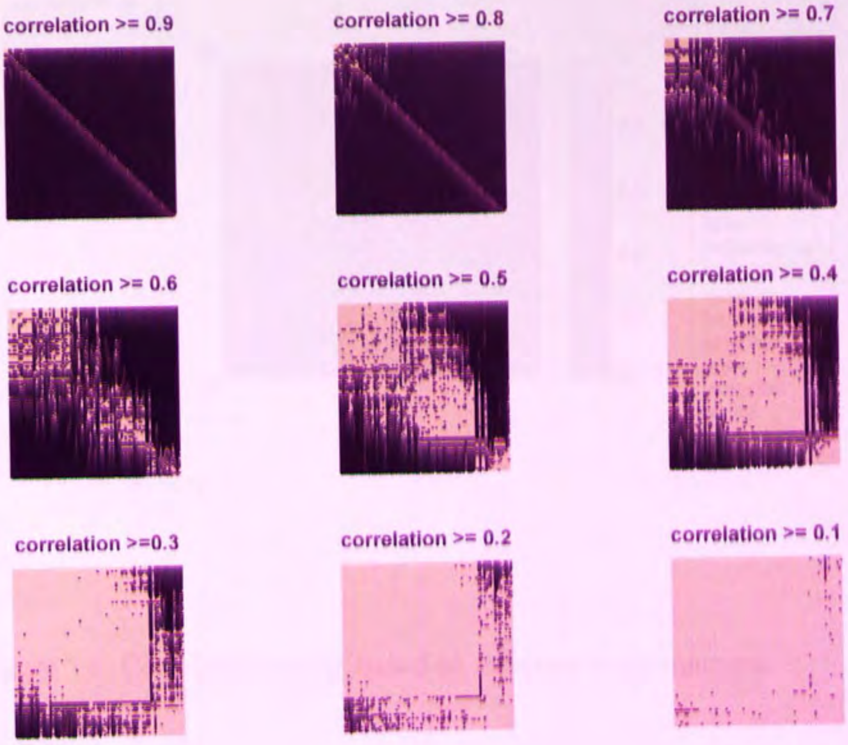


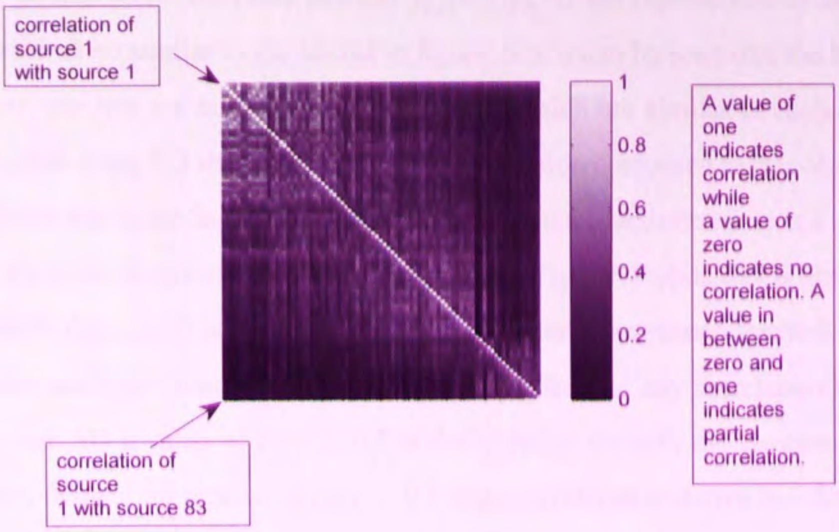
Figure 5.5: Correlation matrices based on rank order of the spectra of the sources for different thresholds on the correlation factors.

from top right to bottom left are noticed. These bars divide the plot into sections. $f_s(w)$ in each section are correlated to each other by a factor of 0.7 or above indicating that the **s** in the section have similar spectra. Therefore by considering a $corr_{\%} = 0.7$ it is possible to obtain clusters having individual characteristics.

Structure order matrix clustering:

Since $\mathbf{r}_{f_a}(i)$ is obtained from the comparison of the $f_{s_a}(i)$ with all the other samples in $f_{s_a}(w)$ taken together, we hypothesised that the correlation between **SO** of each pair of $f_{s_a}(w), f_{s_b}(w)$ where $a, b \in \{1, 83\}$ of the 83 estimated **s** might provide better classification than the correlation between $\mathbf{r}_f$ of each pair of $f_{s_a}(w), f_{s_b}(w)$. Figure 5.6 is the correlation matrix obtained by comparing **SO** of each pair of $f_{s_a}(w)$, $f_{s_b}(i)$ of the 83 sources. This figure is very similar to the plot shown in figure 5.4.

The subplots in figure 5.7 show the correlation matrices obtained for $f_s(w)$ of each of the estimated 83 **s** against $f_s(w)$ of every other **s** using the **SO** of each **s**. The value of



correlation of source 1 with source 1

correlation of source 1 with source 83

A value of one indicates correlation while a value of zero indicates no correlation. A value in between zero and one indicates partial correlation.

Figure 5.6: Correlation matrix based on structure order matrices.

corr is printed above each plot.

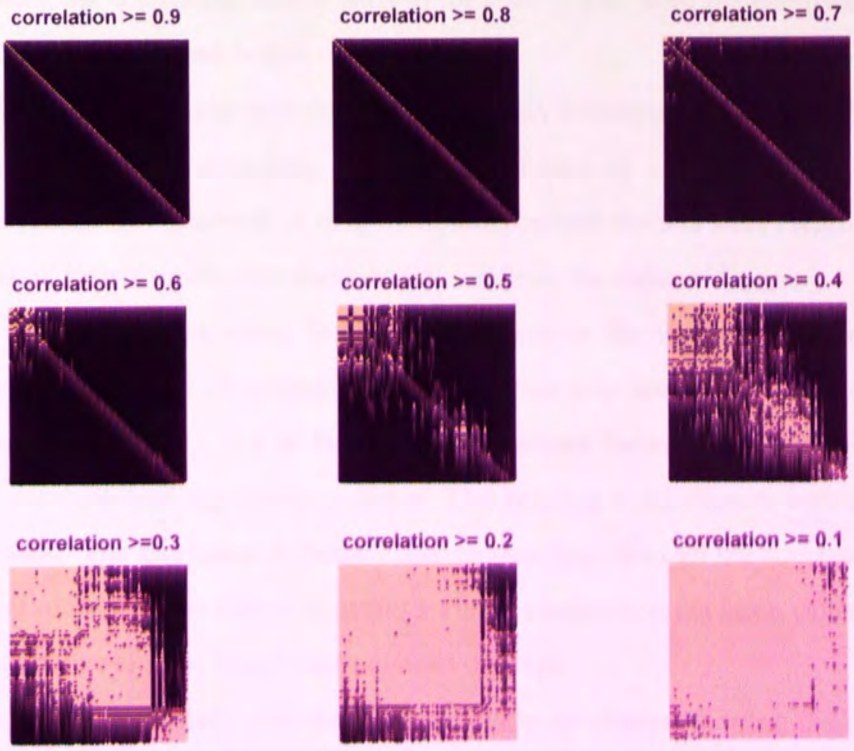


Figure 5.7: Correlation matrices based on structure order matrices for different correlation factors.

The pairs of $f_{s_a}(w)$, $f_{s_b}(w)$ for $a, b \in \{1, 83\}$ which are correlated $\geq$ *corr* mentioned are represented as one while the other pairs of $f_{s_a}(w)$, $f_{s_b}(i)$ are represented as zero as already mentioned. Also similar to the results in figure 5.5, it can be seen that the higher the value of $corr$, the less the number of pairs of $\mathbf{s}_a$, $\mathbf{s}_b$ which are similar to each other. The correlation plots using **SO** show a lower level of correlation compared to that obtained using the $\mathbf{r}_f$. This is due to the fact which has already been mentioned above. $\mathbf{r}_f$ is a vector obtained using the sum of comparison of $f_s(i)$ with all $f_s(j)$; $j \neq i$ while **SO** is obtained by comparing each $f_s(i)$, $f_s(j)$ where $i, j \in \{1, 83\}$. As already mentioned for rank order clustering definite and individual clusters can be obtained. But the any correlation factor required in the case **SO** a value of $corr \geq 0.5$ is sufficient to identify the clusters. The correlation matrix defined for a value of $corr \geq 0.5$ was considered to derive the clusters.

5.4 Comparison of Clustering Using the Euclidean Distance Measure and Ordinal Analysis

A clustering method requires a distance and an algorithm. Some of the results of the clustering method using the correlation matrix derived for $corr \geq 0.5$, from the comparison of **SO** of the 83 sources are shown below.

In order to derive a comparison with the most commonly known and used method of Euclidean distance measure of separating different sets of data $d_{ij}^2 = (\mathbf{s}_i - \mathbf{s}_j)(\mathbf{s}_i - \mathbf{s}_j)'$, the following procedure was followed. A total of 23 independent clusters were identified using the correlation matrix mentioned above and this defined the value of N_{Clust}.

In the case of the clustering using the Euclidean distance the Euclidean distance measure obtained for each pair of spectra of the 83 sources was normalised. Sources whose Euclidean distance was < 0.5 of the maximum distance between all pairs of $\mathbf{s}_i$, $\mathbf{s}_j$; $i, j \in \{1, 83\}$ were grouped together in a cluster. This resulted in 83 clusters with one source in one cluster. The Euclidean distance measure thus identified all the sources as being independent of each other. Hence to obtain a viable comparison the same value of N_{Clust} was used to derive clusters using the Euclidean distance.

Of the 23 independent clusters identified independently by clustering using the Euclidean distance measure and the structure order matrix method, the clusters which are similar to each other in characteristics (one identified using the Euclidean distance and the other using the structure order matrix method) are shown in figures 5.8 and 5.9. Of

s_i $i \in \{1, 83\}$ estimated in no particular order of independence or information content, the Euclidean distance measure clustered sources s_{67}, s_{68}, s_{70} and s_{71} to have a similar spectra and thus belonging to one cluster. The ordinal analysis method (based on the correlation of **SO**) grouped sources s_{68}, s_{70}, s_{71}, s_{78} and s_{80} as being similar and of one group. It can be observed that while the Euclidean distance grouped the sources based on the similarity in the power of the spectra of the sources across the frequency scale, the ordinal analysis method grouped the sources based on the power content of the sources.

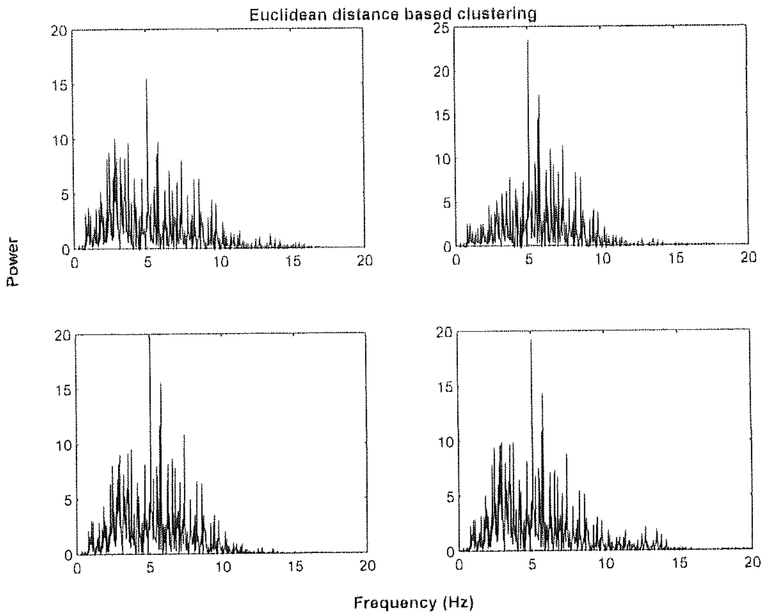


Figure 5.8: Example of a cluster of band limited sources - Euclidean distance measure.

Figure 5.10 shows the spectra of two sources s_{77}, s_{79} which have been grouped in one cluster by the ordinal analysis method but s_{77} and s_{79} were grouped as being independent and in separate clusters by the Euclidean distance method. As already mentioned the two sources s_{77} and s_{79} are different in terms of the power content at each step of the frequency scale but the total power content of the two sources is nearly equal leading to the classification as similar by the ordinal analysis method.

The Euclidean distance measure fails to differentiate between two similar signals shifted slightly or as shown in the example in our experiments (containing the same bandwidth and total power but vary slightly in the power content at each frequency). The Euclidean distance measure also fails in that it does not give any differentiating factor to identify individual clusters. These shortcomings in the Euclidean distance measures can be overcome by using the ordinal analysis methods which might be useful for certain

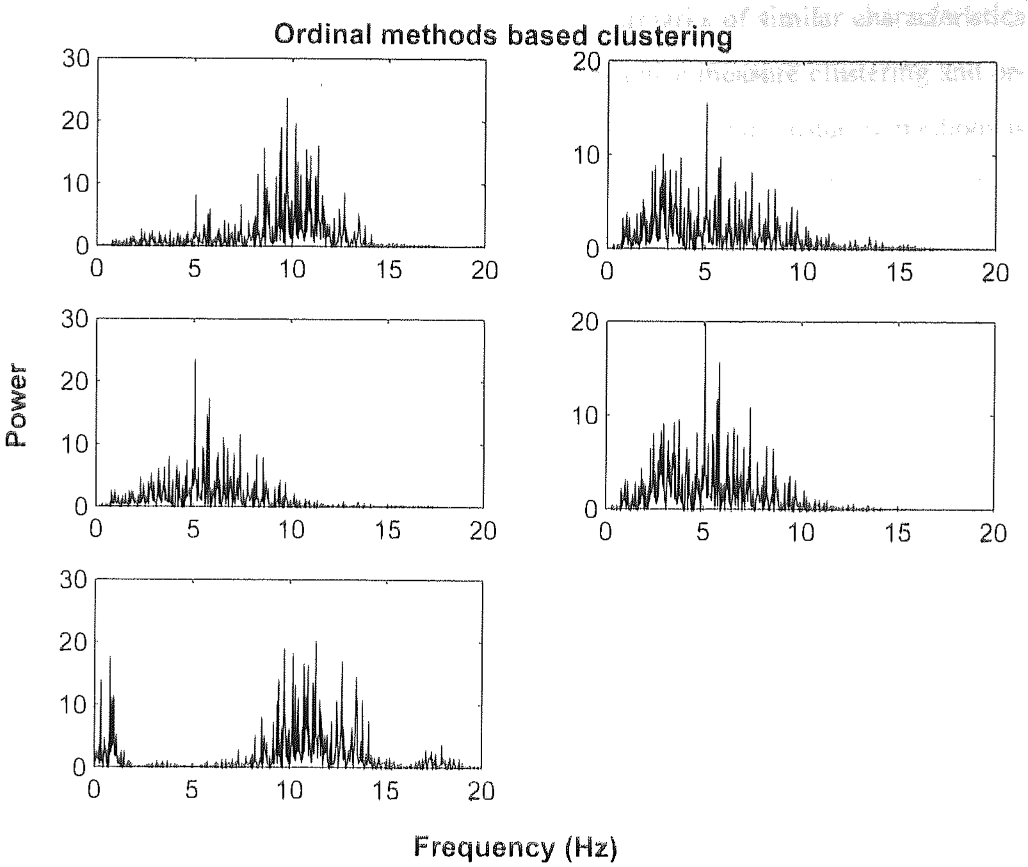


Figure 5.9: Example of a cluster of band limited sources - ordinal analysis method.

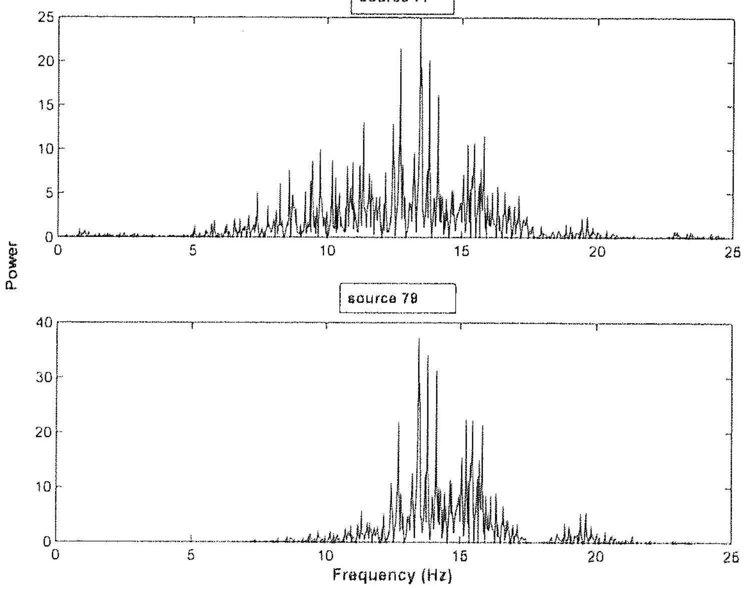


Figure 5.10: Example of a wide band cluster using Euclidean distance measure.

applications. Hence in the embedding of multiple watermarks of similar characteristics the clustering based on the two methods, Euclidean distance measure clustering and ordinal analysis method of clustering is used. The results of the two clustering methods is compared by calculating the bit error rate, *dist* for each of the four watermarks retrieved at the decoder.

5.5 Choice of Cluster

Figures 5.11, 5.12 show the spectra of the four sources used to embed the four watermarks $\mathbf{m}_{sim_1}$, $\mathbf{m}_{sim_2}$, $\mathbf{m}_{sim_3}$, $\mathbf{m}_{sim_4}$ for the two methods heirarchical clustering and cluster based on ordinal analysis. Each watermark represents the identity of the individual who has accessed the medical record. As mentioned previously all the watermarks require the same level of security and robustness, hence four sources whose spectra are similar are required. The four sources whose spectra are shown in figure 5.11 are clustered together by the Euclidean distance based clustering algorithm. Sources 71, 70, 68 and 67 are used to embed the four watermarks $\mathbf{m}_{sim_1}$, $\mathbf{m}_{sim_2}$, $\mathbf{m}_{sim_3}$, $\mathbf{m}_{sim_4}$ sequentially. Similarly the four sources shown in figure 5.12 are clustered together using the ordinal analysis based clustering algorithm. Sources 80, 70, 68 and 78 are used to embed the four watermarks $\mathbf{m}_{sim_1}$, $\mathbf{m}_{sim_2}$, $\mathbf{m}_{sim_3}$, $\mathbf{m}_{sim_4}$ sequentially.

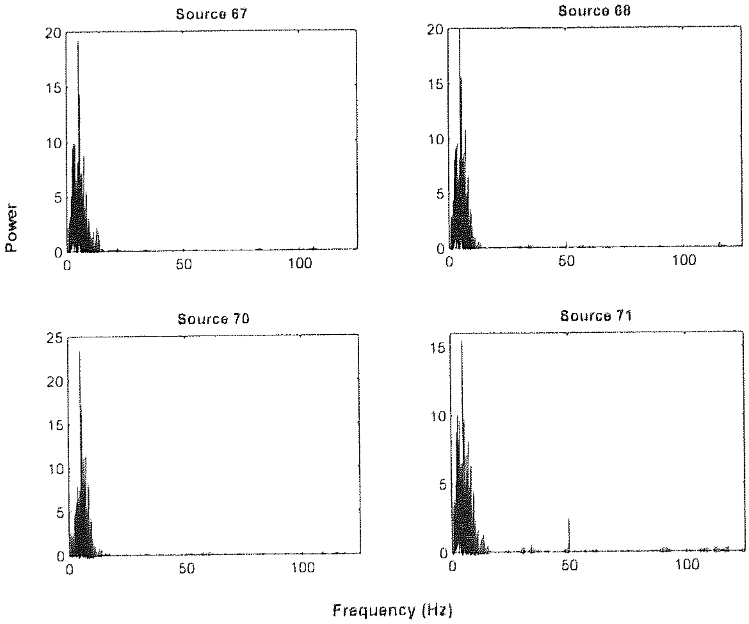


Figure 5.11: Four similar source spectra clustered using the Euclidean distance measure.

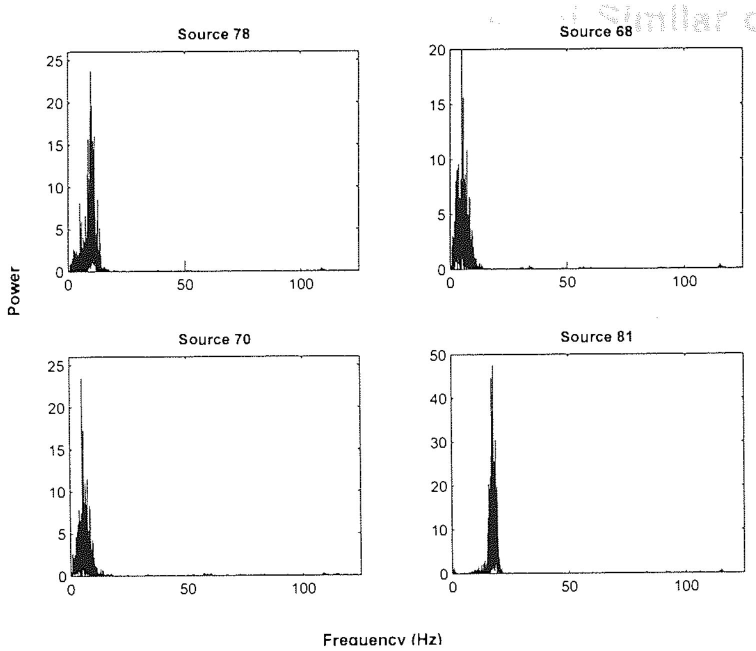


Figure 5.12: Four similar source spectra clustered using the ordinal analysis method.

Figure 5.1 depicts six sources which have distinct spectra. This shows that the delay embedding method can estimate **s** which are distinct and relatively independent from one-dimensional **c**.

For a watermark to be robust as mentioned in the previous chapters, it must be embedded in the informative domain of **c** which typically means low to middle frequency components of **c**. Therefore the cluster chosen should contain sources **s** which are representative of this spectra. One major problem with embedding multiple watermarks of similar properties is that the transformation of **c** into S might not result in a cluster containing the same number of sources, **s** as the number of watermarks. Hence the number of watermarks embedded is upper bounded by the cluster chosen and the number of distinct sources **s** in the cluster. Alternately in the case of the number of watermarks being fixed, a cluster can be chosen with the required number of sources **s**. This cluster can contain sources **s** containing higher or lower spectra thus affecting the characteristics of the embedded watermarks. Hence the practical implementation of our proposed method will entail compromise.

5.6 Embedding Multiple Watermarks of Similar characteristics

The cover work $\mathbf{c}$ is to be embedded with different watermarks at different instants of time. Every successive watermark is embedded in the currently available $\hat{\mathbf{c}}$. Neither the choice of the cluster nor the identity of source/sources previously watermarked is stored. Hence when embedding a new watermark, the selection of an unwatermarked source and also the identical cluster used for embedding previous watermarks is necessary. Embedding of a watermark in a one-dimensional $\mathbf{c}$ using the ICA method discussed in chapters 3 and 4 is used for the experiments in this chapter. Let $\hat{\mathbf{c}}_{sim_i}$ represent a compressed watermarked cover and the value of i represent the number of watermarks embedded.

- Consider four watermarks $\mathbf{m}_{sim_1}$, $\mathbf{m}_{sim_2}$, $\mathbf{m}_{sim_3}$, $\mathbf{m}_{sim_4}$ that are to be embedded in $\mathbf{c}$ sequentially at different instants of time.

- For watermark 1:

  The delay embedding method to obtain $\mathbf{X}$ from $\mathbf{c}$ is used.

$$\mathbf{c} \rightarrow \mathbf{X}. \tag{5.9}$$

$$[\mathbf{S}, \mathbf{W}] \overset{ICA}{\longleftarrow} \mathbf{X} \tag{5.10}$$

  Cluster the sources using Euclidean distance.

  Select the cluster with the required properties.

  Embed $\mathbf{m}_{sim_1}$ in one of the sources from the selected cluster to obtain $\tilde{\mathbf{S}}$.

$$inv(\mathbf{W})\tilde{\mathbf{S}} \rightarrow \tilde{\mathbf{X}} \tag{5.11}$$

  The reconstruction method II described in chapter 2 is used to obtain $\tilde{\mathbf{c}}$ from $\tilde{\mathbf{X}}$.

$$\tilde{\mathbf{X}} \rightarrow \tilde{\mathbf{c}}. \tag{5.12}$$

$$\tilde{\mathbf{c}} + \eta \rightarrow \hat{\mathbf{c}}_{sim_1}. \tag{5.13}$$

- For watermarks 2, 3 and 4; i = 2, 3 and 4:

$$\hat{\mathbf{c}}_{sim_i} \rightarrow \hat{\mathbf{X}}_{sim_i}. \tag{5.14}$$

$$\hat{\mathbf{S}}_{sim_i} = \mathbf{W}\hat{\mathbf{X}}_{sim_i} \tag{5.15}$$

Cluster the sources using Euclidean distance.

Select the cluster with the required properties.

The selection of the unwatermarked source from the chosen cluster was based on the experiments discussed in chapter 3, section 3.3. An estimate of $\breve{\mathbf{k}}_u$ for each source in the cluster is obtained.

$$\breve{\mathbf{k}}_u = [j : \frac{\hat{C}_j}{\breve{\delta}} \leq \varepsilon]. \tag{5.16}$$

This step is required since the specific cluster and source chosen are not stored. The maximum value of $\breve{\mathbf{k}}_u$ for each source, $\breve{\mathbf{k}}_{max}$ is calculated. The source with the smallest value of $\breve{\mathbf{k}}_{max}$ is used to embed the next watermark. Let the watermarked sources with the i watermarks be represented as $\tilde{\mathbf{S}}_{sim_i}$.

$$inv(\mathbf{W})\tilde{\mathbf{S}}_{sim_i} \rightarrow \tilde{\mathbf{X}}_{sim_i} \tag{5.17}$$

The reconstruction method II described in chapter 2 is used to obtain $\tilde{\mathbf{c}}_{sim_i}$ from $\tilde{\mathbf{X}}_{sim_i}$.

$$\tilde{\mathbf{X}}_{sim_i} \rightarrow \tilde{\mathbf{c}}_{sim_i}. \tag{5.18}$$

$$\tilde{\mathbf{c}}_{sim_i} + \eta \rightarrow \hat{\mathbf{c}}_{sim_i}. \tag{5.19}$$

The experiments conducted are realised on the same clustering of sources for each new watermark. Since the identities of the sources in each cluster is not saved, the clustering of the sources is conducted when each new watermark needs to be embedded. The watermarked cover is subjected to different signal processing attacks such as compression, filtering. The clustering algorithm resulted in the same clusters of sources from the attacked watermarked cover $\hat{\mathbf{c}}$. The same clusters of sources were not obtained from $\hat{\mathbf{c}}$ when it was filtered to contain a frequency composition less than 40% of the spectra of the unwatermarked cover. A filter operation of such magnitude destroys the cover to a large extent rendering it unuseable for any application. Hence such severe signal processing operations on $\tilde{c}$ are not valid and so are ignored.

5.7 Results

Four watermarks $\mathbf{m}_{sim_1}$, $\mathbf{m}_{sim_2}$, $\mathbf{m}_{sim_3}$, $\mathbf{m}_{sim_4}$ are embedded in four sources $\mathbf{s}_{sim_1}$, $\mathbf{s}_{sim_2}$, $\mathbf{s}_{sim_3}$, $\mathbf{s}_{sim_4}$ sequentially. If no attack is applied and $\tilde{\mathbf{c}}_{sim_i}$ is embedded with $\mathbf{m}_{sim_{i+1}}$ sequentially as explained above in section 5.6, all the four watermarks can be recovered with only very slight distortion as shown in figures 5.13, 5.14. This is the benefit of the ICA approach which aims for independent sources.

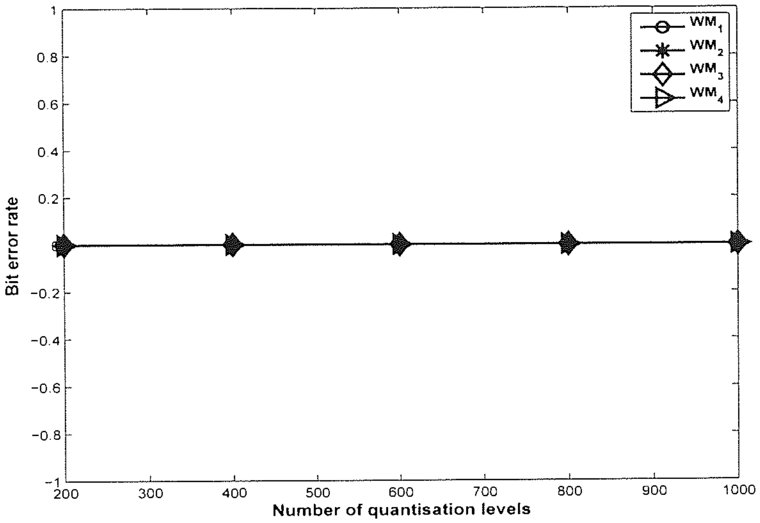


Figure 5.13: Four embedded watermarks in the absence of an attack, Euclidean distance measure.

When the watermarked content is not processed further and $\hat{\mathbf{c}}_{sim_i}$ is embedded with $\mathbf{m}_{sim_{i+1}}$ sequentially as explained above in section 5.6, It can be seen that the watermarks degrade in the sequence they were embedded.

Figures 5.15, 5.17, 5.19, 5.21 depict the error rate in recovering the four watermarks after compression, requantisation, low pass filtering and noise addition respectively for the sources clustered using Euclidean distance measure. Figures 5.16, 5.18, 5.20 5.22 depict the error rate in recovering the four watermarks after compression, requantisation, low pass filtering and noise addition respectively for the sources clustered using Ordinal distance measure. It can seen that the results in the case of the Euclidean distance measure clustering the order of the sequence cannot be retrieved in the absence of an attack as the bit error rate is zero for all the watermarks. Similarly the four watermarks embedded in the cluster obtained from Ordinal analysis based clustering are also recovered with zero bit error rate (one of the watermarks is recovered with an error of 0.02 which is negligible). In the case of an attack, the watermarks embedded in clusters obtained using both Euclidean based clustering and Ordinal analysis can be recovered in the sequence they were embedded when the bit error rates are above 10% (which is negligible) and less than 50% (the attack on the cover work is large which means the cover work is also

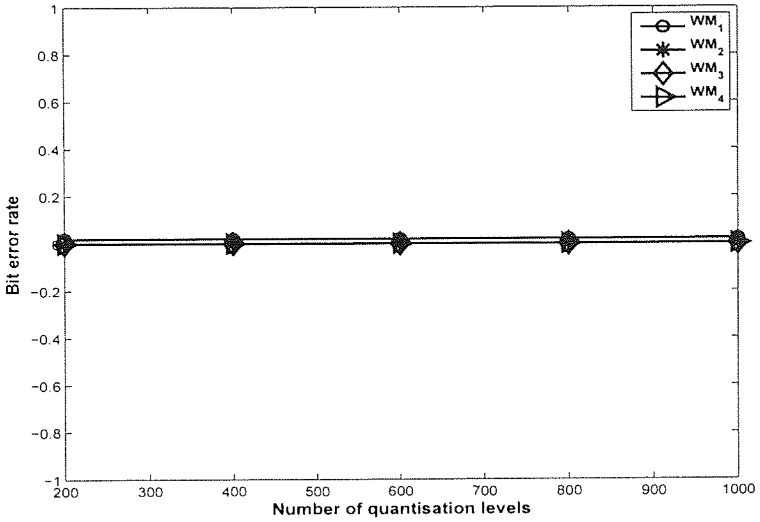


Figure 5.14: Four embedded watermarks in the absence of an attack, ordinal analysis method.

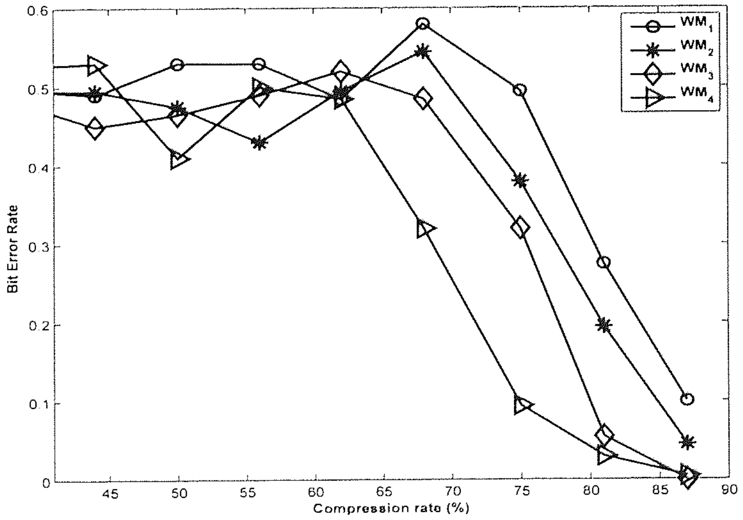


Figure 5.15: Four embedded watermarks recovered after a compression attack, Euclidean distance measure. Note that the order of degradation of the messages follows the sequence of message hiding.

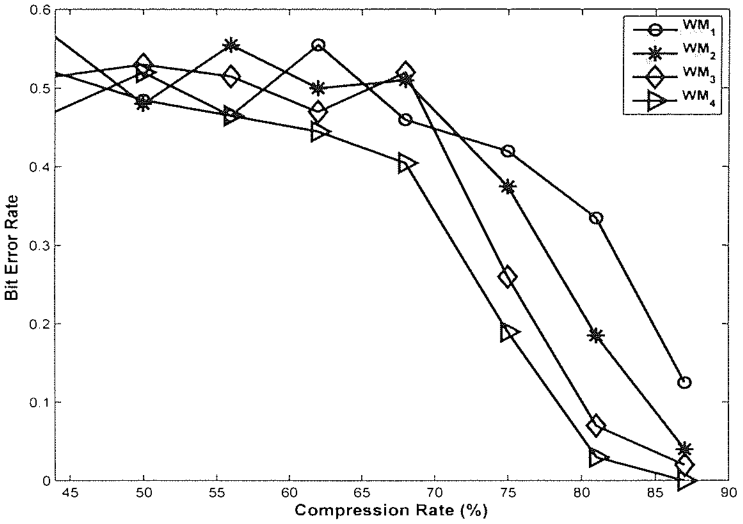


Figure 5.16: Four embedded watermarks after a compression attack, ordinal analysis method. Note that the order of degradation of the messages follows the sequence of message hiding.This shows that the four watermarks can be recovered in the sequence in which they were embedded in the presence of an attack.

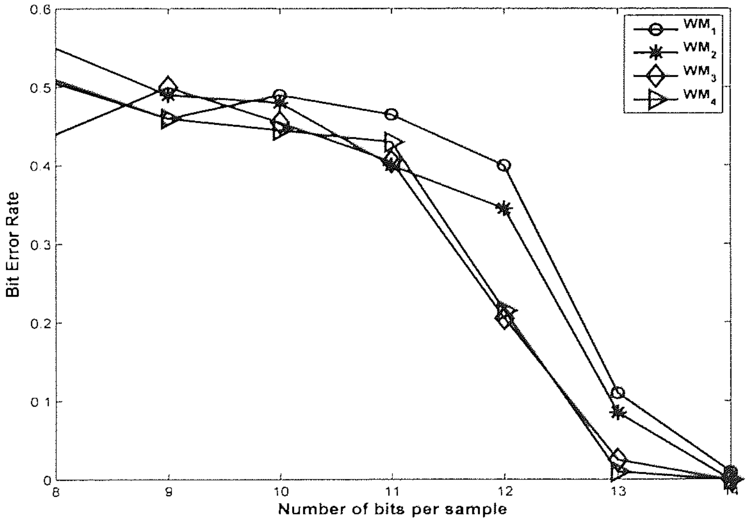


Figure 5.17: Four embedded watermarks recovered after a requantisation attack, Euclidean distance measure. Note that the order of degradation of the messages follows the sequence of message hiding.

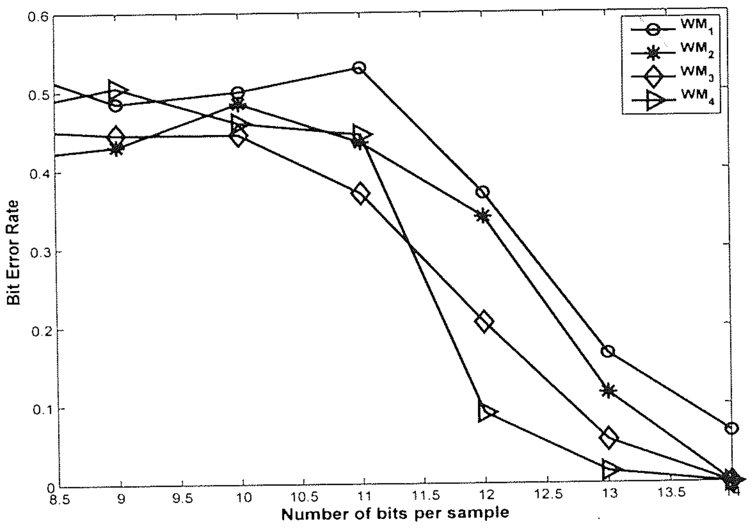


Figure 5.18: Four embedded watermarks after a requantisation attack, ordinal analysis method. Note that the order of degradation of the messages follows the sequence of message hiding.

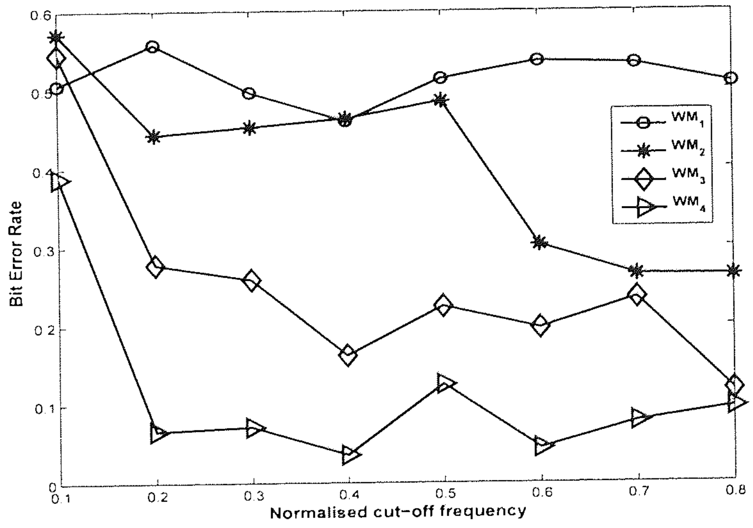


Figure 5.19: Four embedded watermarks recovered after a low pass filtering attack, Euclidean distance measure. Note that the order of degradation of the messages follows the sequence of message hiding.

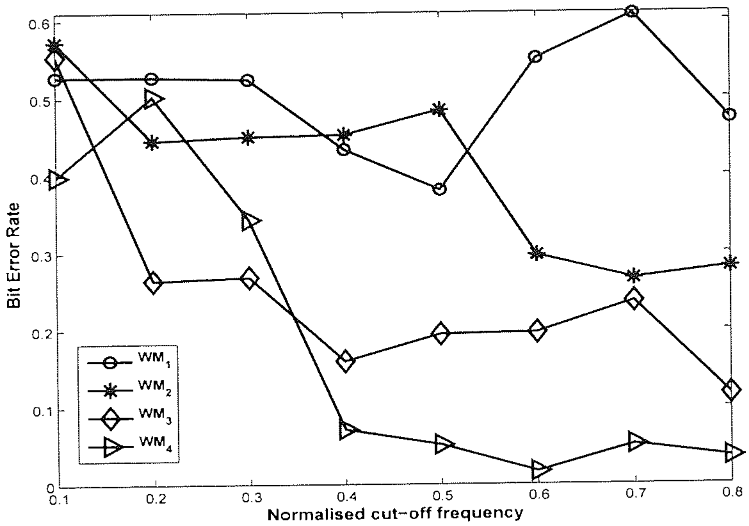


Figure 5.20: Four embedded watermarks in the absence of a low pass filtering attack, ordinal analysis method. Note that the order of degradation of the messages follows the sequence of message hiding.

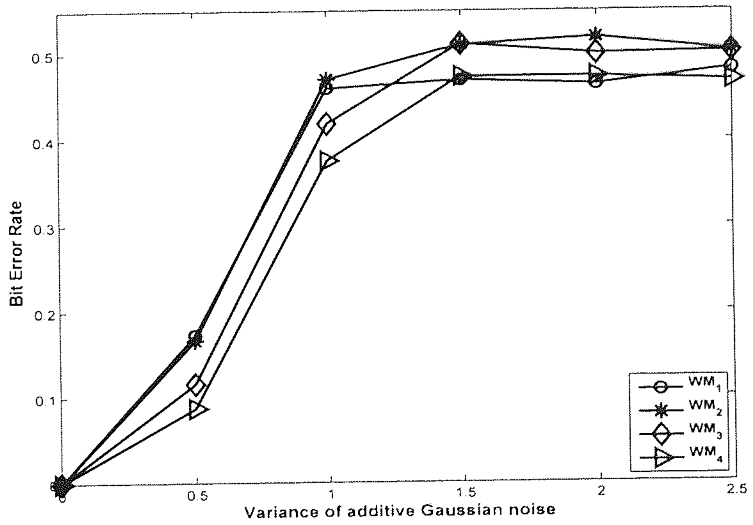


Figure 5.21: Four embedded watermarks recovered after an additive noise attack, Euclidean distance measure. Note that the order of degradation of the messages follows the sequence of message hiding.

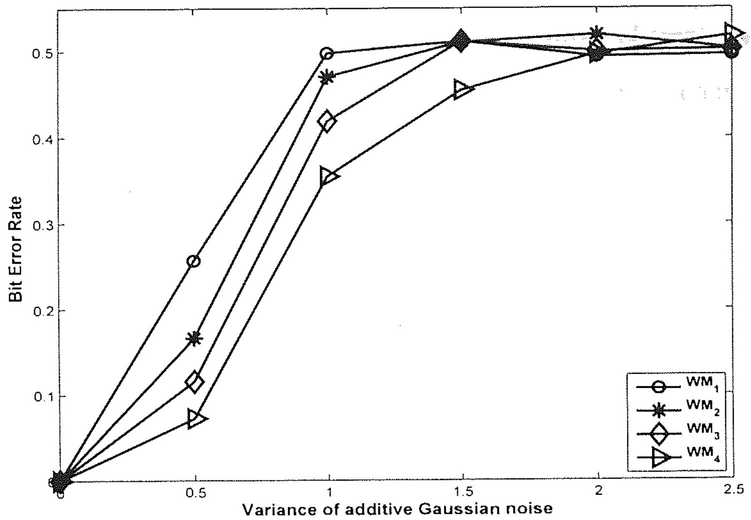


Figure 5.22: Four embedded watermarks after an additive noise attack, ordinal analysis method. Note that the order of degradation of the messages follows the sequence of message hiding.

destroyed).

The experiments were conducted on 22 EEG signals taken from different data sets (EEG from patients undergoing seizures and EEG with no abnormal activity). As discussed in chapter 4, section 4.7, since in our experiments the recording taken at each node (one-dimensional EEG) is considered as an individual cover text, the correlations or abnormalities did not affect the actual watermark embedding/decoding method, or results.

5.8 Conclusion

It was shown that by using the ICA method, numerous sources of equal robustness can be obtained. By embedding multiple watermarks in these sources not only were all the embedded watermarks recovered but by applying three different attacks (compression, low pass filtering, addition of noise), the order of the embedded watermarks was also determined. This is particularly useful if a record of the people who have accessed a medical record is to be maintained securely.

The multiple watermarks of similar characteristics were embedded into sources grouped using two different methods of classification, Euclidean distance measure and ordinal analysis. From the results obtained the group of sources clustered using ordinal analysis provide a stable sequencing of the watermarks both in the presence and absence of an attack. Hence we conclude that ordinal analysis based methods provide a better classification of sources for embedding multiple watermarks of similar characteristics.

In the next chapter the sensitivity of the ICA method to the input data is studied. It will be shown how this sensitivity of the ICA can be succesfully used to secure the embedded data in watermarking applications.

6 SENSITIVITY OF THE ICA

CONTENTS

This thesis has used the properties of an independent component analysis as a basis for watermarking. Amongst the characteristics noted, was a sensitivity to signal reconsruction due to small perturbations in the sources because of data embedding. In this chapter a demonstration of how this sensitivity of the ICA, can be successfully used to secure the embedded data in watermarking applications, is provided.

6.1 Introduction

Blind Source Separation (BSS) is a well known signal processing technique used in analysing a mixed set of data generated from multiple sources. The application of one of the popular BSS techniques, the ICA method, for watermarking applications has been presented in [16, 46]. The sensitivity of the ICA method to the input data has not been explicitly defined as an advantage in watermarking applications in work using the ICA for watermarking. The sensitivity of the ICA and its application in the watermarking scenario was first mentioned in [97]. An analysis of this sensitivity was conducted and presented in [55]. In the following sections the ICA and the sensitivity issue is explained in more detail.

The ICA [42] method for blind source separation extracts a set of basis vectors $\mathbf{W} = \{\mathbf{w}_i; i = 1, \ldots, n\}$ from the given set of mixed observations $\mathbf{X}$. The set of observation vectors $\mathbf{X}$ is projected onto $\mathbf{W}$. These projections being statistically independent, represent the underlying sources $\mathbf{S}$ which generated the mixed observations,

$$\mathbf{WX} \rightarrow \mathbf{S}. \tag{6.1}$$

The ICA method is stable in extracting the underlying sources for slight perturbations in the input data but it is sensitive in that slight changes in the input set of mixed observations above a threshold result in differing values of the independent components. The final projections or the estimated sources $\mathbf{S}$ from the perturbed set of observations $\mathbf{X}$ are altered such that the embedded information is lost but the physical characteristics of the sources are not changed.

Embedding a message $\mathbf{m}$ into a cover work $\mathbf{c}$ distorts $\mathbf{c}$ to a degree that is perceptually invisible. Hence the watermark is nothing but the distortion to $\mathbf{c}$, $\mathcal{D}_{Emb}$.

$$\mathbf{c} + \mathcal{D}_{Emb} \rightarrow \tilde{\mathbf{c}}. \tag{6.2}$$

The watermarked cover $\tilde{\mathbf{c}}$ is usually subjected to intentional/malicious signal processing attacks, η during transmission thus increasing $\mathcal{D}_{Emb}$.

$$\tilde{\mathbf{c}}+\eta \rightarrow \hat{\mathbf{c}}. \tag{6.3}$$

As already defined in chapter 2, let the distortion to $\tilde{\mathbf{c}}$ be represented as $\varepsilon_{\mathbf{c}}$,

$$\varepsilon_{\mathbf{c}} = \mathcal{D}_{Emb}+\eta. \tag{6.4}$$

Retrieving an estimate of $\mathbf{m}$, $\hat{\mathbf{m}}$ from $\hat{\mathbf{c}}$ therefore depends on $\varepsilon_{\mathbf{c}}$. The larger the value of $\varepsilon_{\mathbf{c}}$ the less the probability of recovering the embedded $\mathbf{m}$. Let ζ define a threshold for η. The value of ζ is not fixed and varies with the type of $\mathbf{c}$, the watermark embedding method $\mathcal{F}$, the type of attack(s) η and the strength of the attack(s).

$$\mathcal{H}_{\varepsilon_{\mathbf{c}}} = \begin{cases} 1, & \text{if } \varepsilon_{\mathbf{c}} \leq \zeta, \\ 0, & \text{if } \varepsilon_{\mathbf{c}} \geq \zeta. \end{cases} \tag{6.5}$$

$\mathcal{H}_{\varepsilon_{\mathbf{c}}}$ equal to one indicates that $\mathbf{m}$ is recovered with no error and $\mathcal{H}_{\varepsilon_{\mathbf{c}}}$ equal to zero indicates that $\hat{\mathbf{m}} \neq \mathbf{m}$. .

In most of the ICA based watermarking techniques the set of mixed observations $\mathbf{X}$ are taken from $\mathbf{c}$, and $\mathbf{W}$ is used as the key to obtain a transformed set of vectors from X, S. S is used as the embedding space for $\mathbf{m}$. $\mathcal{F}$ defining the embedding function, $\mathbf{k}$ samples of S are selected as the probable embedding locations of $\mathbf{m}$. The selection of $\mathbf{k}$ is based on the application of $\mathbf{c}$ and $\mathbf{m}$.

$$\mathcal{F}(\mathrm{S}(\mathbf{k}),\mathbf{m}) \rightarrow \tilde{\mathrm{S}}. \tag{6.6}$$

At the decoder the modified source $\tilde{\mathrm{S}}+\varepsilon_{\mathbf{c}}$ is used to retrieve an estimate of $\mathbf{m}$, $\hat{\mathbf{m}}$.

This chapter demonstrates how small perturbations to $\mathbf{X}$ can result in large perturbations in the estimation of $\mathbf{W}$. It is shown how the relationship between the perturbations in $\mathbf{X}$ and $\mathbf{W}$ provides a security mechanism for watermarking applications.

6.2 Sensitivity of the ICA

Given observations $\mathbf{X}$, the ICA algorithm estimates the separating matrix $\mathbf{W}$ and the probable statistically independent sources S. Assume $\check{\mathrm{S}}$ is a set of statistically independent sources which are not observable,

$$\check{\mathrm{S}}_{p\times \mathrm{N}_c} = [\check{\mathrm{s}}_1,\ldots,\check{\mathrm{s}}_p], \tag{6.7}$$

$$\boldsymbol{Pr}(\cap_{i=1}^{p}\check{\mathbf{S}}_i) = \prod_{i=1}^{p}\boldsymbol{Pr}(\check{\mathbf{S}}_i). \tag{6.8}$$

Let $\check{\mathbf{A}} \in \mathbb{R}^{p\times p}$, such that

$$\mathbf{X} = \check{\mathbf{A}}\check{\mathbf{S}}, \tag{6.9}$$

where $\check{\mathbf{A}}$ is the unknown mixing matrix.

6.2.1 Sensitivity problem

As already mentioned $\mathbf{W}$ is derived from $\mathbf{X}$. The estimated $\mathbf{W}$ by the ICA algorithm is theoretically the inverse of the unknown $\check{\mathbf{A}}$. Let ξ represent the threshold of the perturbation of $\mathbf{X}$.

$$\mathbf{X} + \xi \rightarrow \bar{\mathbf{X}}. \tag{6.10}$$

$$\mathbf{X} \overset{ICA}{\rightarrow} \mathbf{WS}. \tag{6.11}$$

$$\bar{\mathbf{X}} \overset{ICA}{\rightarrow} \bar{\mathbf{W}}\bar{\mathbf{S}}. \tag{6.12}$$

It was shown in [97] that if $\bar{\mathbf{S}} - \mathbf{S} = \varepsilon_{\mathbf{c}}$ and

$$\mathbf{D}_{Emb} + \varepsilon_{\mathbf{c}} \gg \zeta. \tag{6.13}$$

then $\mathcal{H}_{\varepsilon_{\mathbf{c}}} = 0$.

The sensitivity problem in this thesis is therefore defined as 'the problem of defining bounds for ξ which in turn will define the bounds for $\varepsilon_{\mathbf{c}}$ which will affect the decision $\mathcal{H}_{\varepsilon_{\mathbf{c}}}$'.

The eigenvalues of a fixed matrix give an understanding of the underlying structure of the matrix, but the eigenvalues of sample covariance matrices give information about the underlying distribution. This concept of the eigenvalues and the covariance matrices is exploited in data analysis methods such as the PCA and ICA. Hence the covariance of $\mathbf{X}$ is the fundamental starting point used by the ICA to estimate $\mathbf{W}$ and we rely on the condition number of the covariance to provide an understanding of the sensitivity issue.

Therefore to illustrate the effect of the perturbations on the retrieval of the embedded watermark the following experiment was conducted.

1. A one-dimensional EEG signal $\mathbf{c}$ of twenty seconds is considered as the cover work. The EEG is transformed into a matrix of observation vectors, $\mathbf{X}$ using the dynamical embedding method described in chapter 2.

2. The ICA approach is used to retrieve probable independent sources $\mathbf{S}$ and a separating matrix $\mathbf{W}$. $\mathbf{W}$ is used as one of the keys to retrieve the watermark.

3. One of the sources, $\mathbf{s}_{wm}$ thus obtained is watermarked using the QIM method of message embedding. Let $\tilde{\mathbf{s}}_{wm}$ represent the watermarked source. $\tilde{\mathbf{S}}$ represents the watermarked source matrix.

4. The watermarked EEG $\tilde{\mathbf{c}}$ is reconstructed from $\tilde{\mathbf{X}}$ which is obtained by applying the inverse of the separating matrix, $\mathbf{A}$ to $\tilde{\mathbf{S}}$.

5. To estimate the sensitivity of the ICA method $\mathbf{c}$ is perturbed by a zero mean random noise signal, η to obtain $\bar{\mathbf{c}}$. The variance of the noise signal represents $\varepsilon_{\mathbf{c}}$ and is varied from 0.01 to 0.04.

6. $\bar{\mathbf{c}}_i$ obtained for each value of $\varepsilon_{\mathbf{c}}$ is transformed into a matrix of observation vectors, $\bar{\mathbf{X}}_i$ where i indexes the different noise levels.

7. The condition number of the covariance of $\mathbf{X}$ and each $\bar{\mathbf{X}}_i$ is calculated.

8. The norm of the difference between $\mathbf{X}$ and its perturbed version $\bar{\mathbf{X}}_i$, ξ is calculated.

9. The ICA is applied to each $\bar{\mathbf{X}}_i$ to obtain $\bar{\mathbf{S}}_i$ and $\bar{\mathbf{W}}_i$. The ICA is initialised using the eigen vectors of the covariance of $\mathbf{X}$ in order to obtain the same order of the estimated sources.

10. The condition number of the covariance of $\mathbf{W}$ and each $\bar{\mathbf{W}}_i$ is calculated.

11. The norm of the difference between $\mathbf{W}$ and its perturbed version $\bar{\mathbf{W}}_i$ is calculated.

12. An estimate of the sources $\bar{\mathbf{S}}_w$, $\bar{\mathbf{S}}_i$ is obtained by applying $\mathbf{W}$ and each $\bar{\mathbf{W}}_i$ respectively to $\tilde{\mathbf{X}}$. An estimate of the embedded message is retrieved from $\tilde{\mathbf{s}}_{wm}$ of $\bar{\mathbf{S}}_w$ and each $\bar{\mathbf{S}}_i$.

13. The Hamming distance between the original embedded watermark and the estimated watermark is noted.

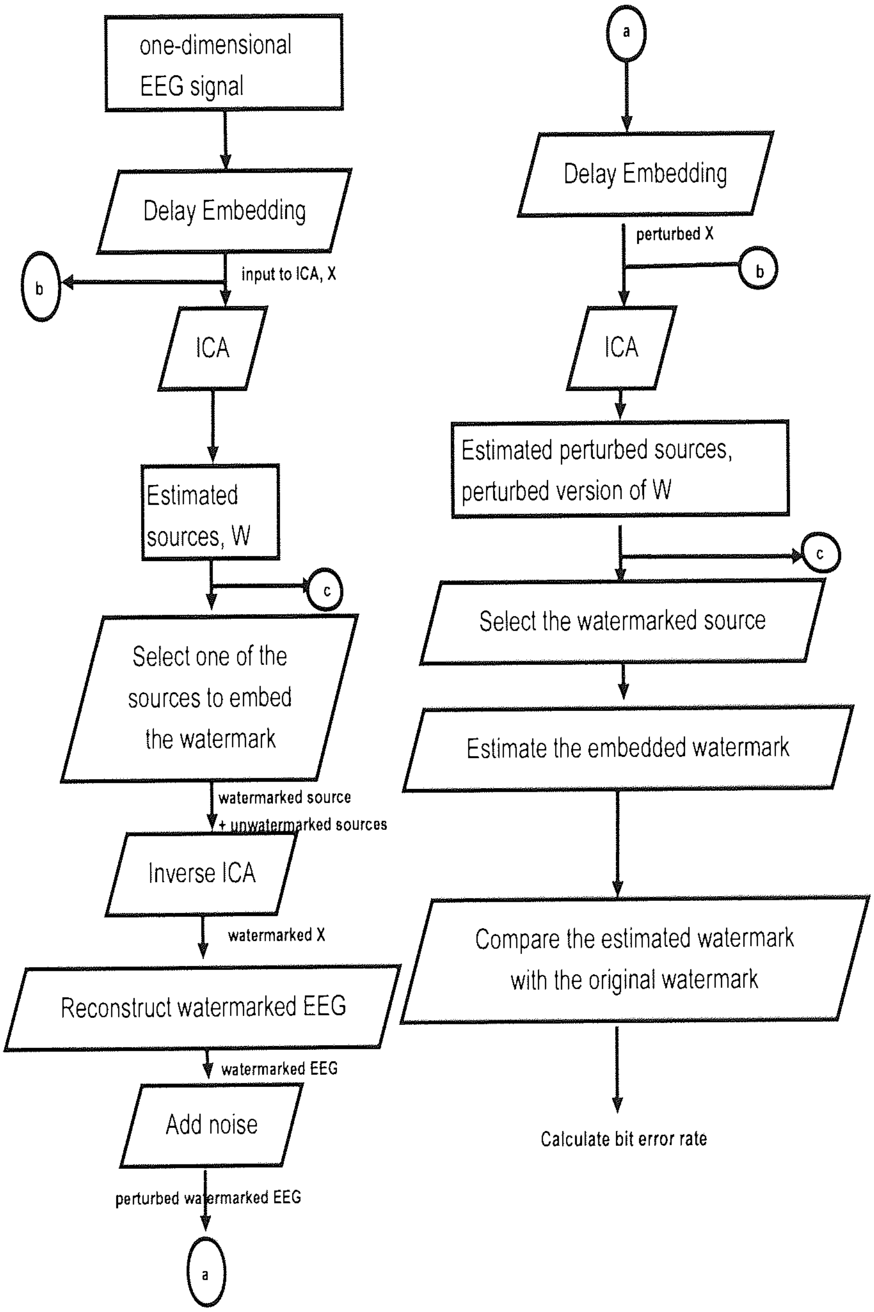

one-dimensional EEG signal
Delay Embedding
input to ICA, X
b
ICA
Estimated sources, W
c
Select one of the sources to embed the watermark
watermarked source + unwatermarked sources
Inverse ICA
watermarked X
Reconstruct watermarked EEG
watermarked EEG
Add noise
perturbed watermarked EEG
a
a
Delay Embedding
perturbed X
b
ICA
Estimated perturbed sources, perturbed version of W
c
Select the watermarked source
Estimate the embedded watermark
Compare the estimated watermark with the original watermark
Calculate bit error rate

Figure 6.1: Schematic diagram illustrating part of the experiment.

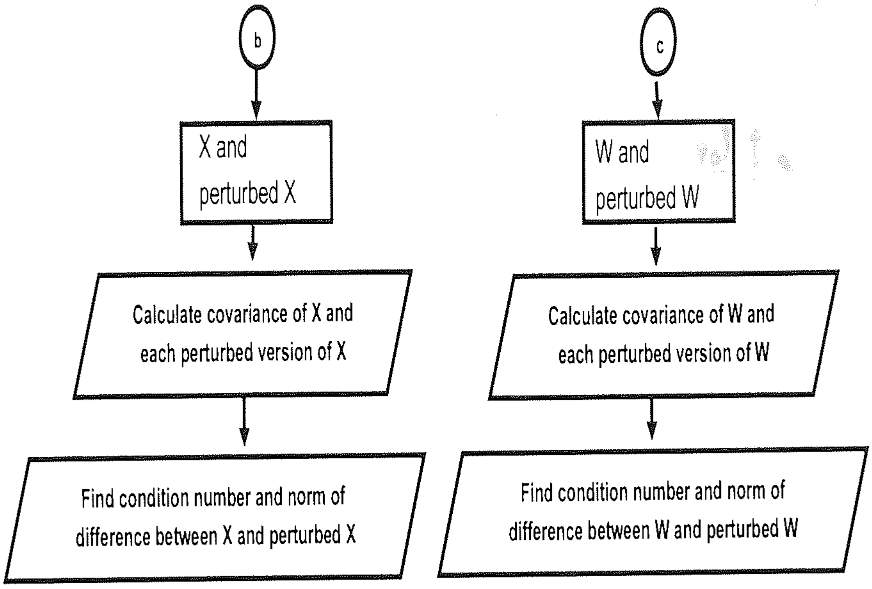

b
c
X and perturbed X
W and perturbed W
Calculate covariance of X and each perturbed version of X
Calculate covariance of W and each perturbed version of W
Find condition number and norm of difference between X and perturbed X
Find condition number and norm of difference between W and perturbed W

Figure 6.2: Schematic diagram illustrating part of the experiment.

6.3 The Application of the Sensitivity of the ICA Method in a Watermarking Application

Figure 6.3 is the condition number of the covariance of the observation matrix, $\mathbf{X}$ and its perturbed versions $\bar{\mathbf{X}}$ plotted against an increasing value of ε_c. As the value of ε_c increases the condition number varies randomly. This is because ε_c is derived from a noise signal which has a normal distribution but the specific samples values of the noise do not conform to any particular order (increasing/decreasing). Hence the effect of the noise signal on each value of the EEG signal varies even though the variance of the noise signal is increasing.

Figure 6.4 is the condition number of the estimated separating matrix $\mathbf{W}$ for $\mathbf{X}$ and the estimated separating matrices $\bar{\mathbf{W}}_i$ for each perturbation of $\mathbf{X}$. The plot of this condition number is similar to the plot of figure 6.3. This shows that any perturbation of the input matrix $\mathbf{X}$ results in an equivalent perturbation of the estimated independent components $\mathbf{W}$. Since $\mathbf{W}$ is used as one of the keys in the retrieval of the embedded message, perturbations of $\mathbf{W}$ will influence the estimate of the embedded message. Figure 6.5 is a plot of the norm of the original observation matrix and its perturbed version for an increasing value of ε_c. This norm represents ξ. It has to be noted that ε_c and ξ have a monotonic relationship. ξ increases with increasing value of ε_c but this phenomenon is not observed in the case of $\mathbf{W}$ derived from $\mathbf{X}$. Figure 6.6 is the norm of the difference between the

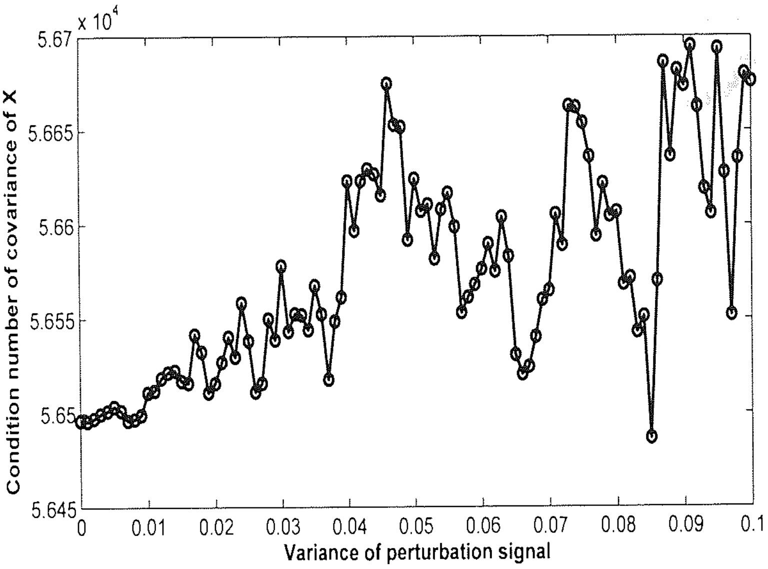


Figure 6.3: Condition number of the covariance of the observation matrix, **X** and its perturbed versions $\bar{\mathbf{X}}$. It is to be noted that the condition number fluctuates as the noise sample fluctuates.

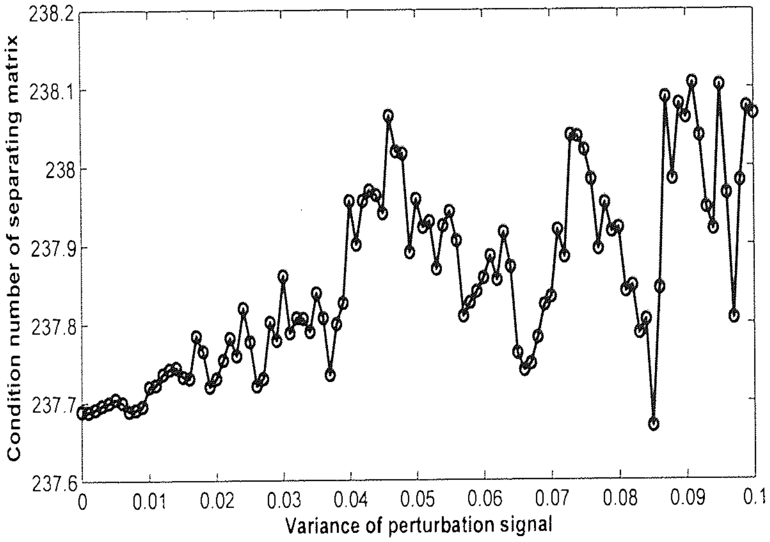


Figure 6.4: Condition number of the covariance of the observation matrix, **W** and its perturbed versions $\bar{\mathbf{W}}$. It is to be noted that the condition number fluctuates. The variation in the condition number of **W** and each $\bar{\mathbf{W}}_i$ follows the changes in the condition number of the respective observation data matrix.

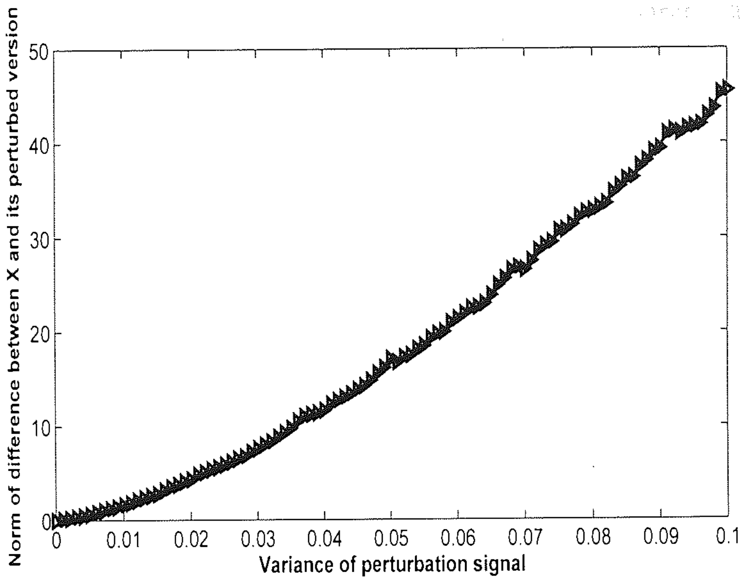


Figure 6.5: Plot of the norm of difference between the observation data and its perturbed version, ξ. As the variation of the perturbation signal, the random noise signal, increases the value of ξ also increases correspondingly and smoothly.

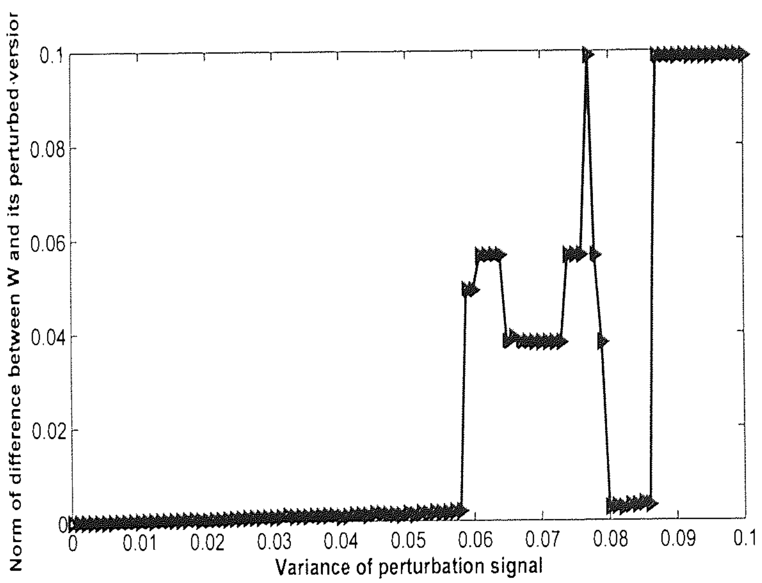


Figure 6.6: Plot of the norm of difference between the separating matrix obtained for the original observation data and the separating matrix obtained for the perturbed version of the observation data. The variation of the norm changes for certain values of the perturbation but for a perturbation signal of large variance the value of the norm is large compared to the initial value.

separating matrix $\mathbf{W}$ and $\bar{\mathbf{W}}_i$ for differing perturbations applied to $\mathbf{X}$. The perturbation of $\mathbf{W}$ is minimal for small values of ε_c but it changes rapidly for larger values of ε_c. An exact value of ε_c for which this change is observed cannot be estimated. This is because ε_c is derived from a noise signal generated randomly at each instant.

Figure 6.7 is the bit error rate in the estimated watermark for each value of ε_c. It can be seen that the bit error rate is zero for small changes in $\mathbf{W}$ but increases to 50% of the embedded message for large changes in $\mathbf{W}$. It has to be noted that the changes in the bit error rate are similar to the changes in the estimated $\mathbf{W}$. In order to demonstrate the relationship between the sensitivity of the ICA, the effect of one of the keys $\mathbf{W}$ and the embedded message the results shown in figures 6.6 and 6.7 are rearranged. The values of the norm of the difference betweeen $\mathbf{W}$ and $\bar{\mathbf{W}}_i$ obtained for different perturbation levels is sorted in the ascending order. Similarly the bit error rate is also ordered. The ordered norms of the difference in the separating matrices and the bit error rate are shown in figure 6.8. It can be observed that when the norm of the difference in $\mathbf{W}$ and $\bar{\mathbf{W}}_i$ increases beyond a threshold the bit error rate increases but for small values of the norm of the difference (below a threshold) the bit error rate is zero.

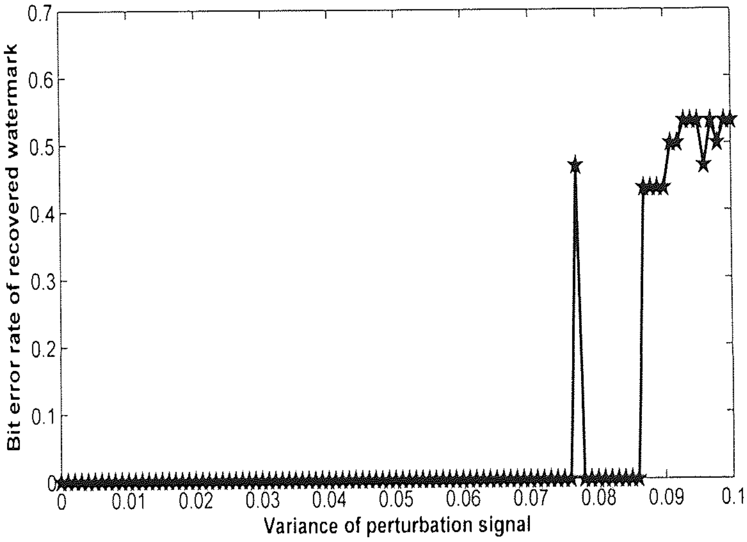


Figure 6.7: Bit error rate of the retrieved watermark. It should be noted that the error rate follows the changes in the norm of the separating matrix. As the norm of the difference between the separating matrix obtained for the original observation data and the separating matrix obtained for the perturbed version of the observation data increases the bit error rate also increases. This shows that the embedded information is lost.

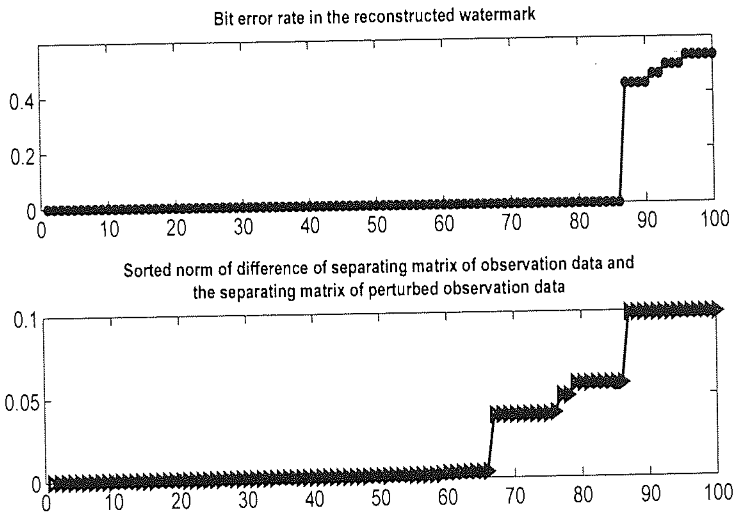


Figure 6.8: Bit error rate of the retrieved watermark sorted in ascending order and sorted norm of the separating matrix shown in figure 6.6. As the norm of the difference between the separating matrix obtained for the original observation data and the separating matrix obtained for the perturbed version of the observation data deviates to a large extent the bit error rate also increases. But for small variations in the norm of the separating matrix the bit error rate is zero. This shows that the embedded information is lost when the input observation data is perturbed beyond a threshold value.

6.4 Conclusion

In this chapter the sensitivity of the ICA to the input data was demonstated. The changes in the estimates of **W** were observed for a simple attack, a random noise signal. It was shown how this sensitivity could provide a solution to the issue of security of the embedded message. The bit error rate in the estimated watermark is high when the perturbation of the observation data increases beyond a threshold. It was also shown how the retrieval of the message is dependent on the usage of the key, **W**. A demonstration of this sensitivity in the watermarking of EPHRs was also given. This work was presented in [58].

7 CONCLUSION

CONTENTS

This thesis provides a partial solution to the problem of patient privacy and security of medical records in the fully connected computerised health care system. In this thesis techniques based on data hiding methods were developed to show that they can provide a parallel security mechanism to the traditional cryptography based ICT methods of security.

The data hiding based security mechanism is shown to resolve some of the problems in the implementation of EPHRs. It is also shown to provide added benefits to the EPHR.

7.1 Embedding Watermarks in One-Dimensional Time Series Data

In chapter 2 as an exercise into understanding the fundamentals of data hiding techniques and their possible application to EPHRs various transform domain based watermarking techniques were implemented. The experiments were conducted on single channel biomedical time series data (for example EEG). The redundancy in a one-dimensional signal is less compared to an image. A watermarking technique which performs efficiently on a one-dimensional signal will perform at least as effectively on a higher dimensional data. These experiments were aimed to provide a solution for the privacy and security issues which work consistently independent of the data type.

In order to provide added value and confidence in the EPHR it was realised that it is important to not only provide a secure channel to store/transmit the personal data but also add extra information. This extra information could possibly be in the form of data related to the genetic composition of the patient, documentation of any diagnoses made, a mechanism to authenticate the medical data, and a log of people (authorised/unauthorised) who have accessed the data.

In chapter 2 the DWT, PCA and ICA based data hiding methods were implemented. It was shown that the various transform domain based watermarking methods provide multiple channels capable of carrying multiple dissimilar characteristics watermarks but only the ICA based data hiding method provided a mechanism to embed multiple similar characteristics watermarks. Though the DFT and DCT based watermarking methods could also be used to embed multiple watermarks, these methods require careful selection of the secret key, **k** for each of the watermarks.

It was also realised that the ICA extracts interesting components and not necessarily independent components from a one-dimensional signal. This resulted in the loss of the embedded data at the embedder itself unless a fixed embedding and reconstruction method was implemented. Hence a new method of embedding information and reconstructing the watermarked one-dimensional signal without loss of the embedded data was designed.

7.2 Security of the Embedded Message, m

The watermark embedding mechanism using the ICA based method of obtaining channels to carry the secret information was based on the QIM method. Every watermark has three main characteristics, imperceptibility, robustness and rate of information. These three characteristics are a trade-off against each other. In the QIM based embedding this trade-off can be controlled by varying the size of the quantisation index δ. Also the locations of the samples $\mathbf{k}$ of the carrier modified to embed the watermark are secret. An illegal attacker who accesses the EPHR has to estimate the true value of δ and $\mathbf{k}$ inorder to retrieve the watermark.

In chapter 3 the claim that scalar QIM and DM-QIM are secure embedding techniques because the value of the δ and $\mathbf{k}$ are unknown to illegal intruders was examined. It was demonstrated how this concept can be thwarted in the case of the DWT based method but upheld in the case of the ICA based technique. The ICA based watermarking method proved to be capable of providing a secure channel to transmit hidden information compared to the DWT based method. It was shown that the a better security mechanism can be designed with the use of extra keys than relying on the embedding technique only as in QIM based embedding methods.

7.3 Enhancing the Value of the EPHR

Embedding a single watermark for copyright or authentication purposes is the norm in multimedia applications. Whereas in this thesis watermarking techniques are applied to an EPHR to not only provide a secure mechanism to store and transmit sensitive patient data but also enhance the value of the EPHR. This is to provide confidence in the use of the EPHR by the public.

In order to enhance the value of the EPHR multiple messages of similar and dissimilar characteristics need to be embedded in the EPHR. In chapter 4 multiple watemarks of dissimilar characteristics are embedded. The multiple watermarks have different requirements of imperceptibility, robustness and rate of information. The first watermark representing patient information requires a high level of robustness, imperceptibility, security but a low data rate. The watermarks representing doctor's notes and identity of clinician/hospital to authenticate the source of the data require moderate to low level of robustness, imperceptibility and security, but they require a higher data rate compared to the first watemark. The last of the watermarks is a fragile watermark used to authenticate the originality of the EPHR and to provide an estimate of the attack η. This watermark requires a low level of robustness, imperceptibility and security. It is fragile in the sense it disintegrates when the EPHR undergoes an attack. Embedding of multiple dissimilar watermarks was conducted using the two transform domain based methods, DWT and ICA.

Though in terms of the trade-off obtained between the two methods of DWT and ICA, the ICA performs slightly better, privacy of the sensitive personal information in the EPHR being the main concern, the ICA is advantageous. This is because the ICA based approach requires extra information in the form of a separating matrix $\mathbf{W}$ to enable correct decoding of the embedded information.

In chapter 5 multiple watermarks of similar characteristics as opposed to dissimilar characteristics watermarks are embedded. If multiple watermarks are used as a log of access of the medical record, the information provided by each watermark is the same. The problem in such watermarking applications lies in deriving the order of the embedded watermarks. A mechanism of sequencing the multiple embedded watermarks at the decoder has been derived in this chapter.

In the ICA method numerous channels of equal robustness can be obtained. By embedding multiple watermarks in these channels not only all the embedded watermarks can be recovered in the absence of an attack but in the presence of an attack, the order of the embedded watermarks can also be determined. This is particularly useful if a record of the people who have accessed a medical record is to be maintained securely.

7.4 Mathematical Explanation of the In Built Security of the ICA

In the final chapter of this thesis the sensitivity of the ICA method to the input data was analysed. An example of how this sensitivity of the ICA can be successfully used to secure the embedded data in watermarking applications was also presented.

The experiments conducted showed that the ICA is a stable algorithm in the sense that for a given set of observations and slightly perturbed versions of the same observations, the resulting estimated sources are structurally the same. It was observed that the ICA was sensitive to slight perturbations of the input observations outside a threshold, in that, the independent components estimated for an observation set and its perturbed version vary. This sensitivity provides an in-built security mechanism. A simple example was used and shown how this sensitivity can be exploited in the watermarking of the EPHRs.

7.5 Future Work

The watermarking of one-dimensional $\mathbf{c}$ and the reconstruction of the watermarked $\mathbf{c}$, $\tilde{\mathbf{c}}$ derived in this thesis is novel. Improvements in the trade off characteristics of the embedded watermarks could be enhanced by deriving a better method of estimating statistically independent sources from a one-dimensional time series signal. One such ICA method based on ordinal analysis 'SWICA' has been published recently in [51]. This method of estimating sources proved to be slow as claimed by the authors. Since the embedding window size, 83 in most of the experiments is large, the SWICA method was unsuccessful in estimating the sources. Hence a comparison of the original results with the new method could not be obtained.

As discussed in chapter one most of the current EPHR standards leave the security protocols to be implemented by third party users. Our recommendation is to build the security protocols within the standard thus maintaining uniformity in both implementation and usage.

We conclude that unless problems involving the security issues and lack of confidence in the use of the EPHR by the general population are not resolved implementation and use of a centralised eHealth system cannot be achieved. This thesis has provided a practical and simple method of partially resolving these issues.

List of publications

1. Watermark-Only Security Attack on DM-QIM Watermarking: Vulnerability to Guided Key Guessing. B. R. Matam and David Lowe. International Journal of Digital Crime and Forensics. Accepted.

2. Watermarking Audio Signals for Copyright Protection Using ICA. B.R.Matam and David Lowe. Advanced Techniques in Multimedia Watermarking: Image, Video and Audio Applications. IGI global publications. Ed. Professor Ali Al-Haj. (book to be published towards the end of 2009).

3. Exploiting Sensitivity of Nonorthogonal Joint Diagonalisation as A Security Mechanism in Steganography. D. Lowe and B.R. Matam. Proc. 16th Int. Conf. Digital Signal Processing, 2009.

4. Watermarking: How Secure is the DM-QIM Watermarking Technique? B. R. Matam and D. Lowe. Proc. 16th Int. Conf. Digital Signal Processing, 2009.

5. Participatory EPHR: A Watermarking Solution. B. R. Matam and D. Lowe. Proc. 10th Int. Work-Conf. Artificial Neural Networks, 2009.

6. Towards Gaining Patient Confidence in, and Acceptance of the Electronic Patient Health Record: A Steganographic Proposal. B.R.Matam, David Lowe. Proc. 3rd International Conference on Computational Intelligence in Medicine and Healthcare, 2007.

7. Standardisation of Medical Formats And Protection of Patient Privacy. B.R.Matam, David Lowe. 3rd Conference on Medical Data Axquisition Standards and Protocols, 2006.

8. Steganography, Biopatterns and Independent Components. David Lowe, B.R.Matam and B.Toch. The 7th IMA International Conference on Mathematics in Signal Processing, 2006.

References

[1] http://www.e-health-insider.com/news/item.cfm?ID=2449.

[2] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.9769.

[3] International guidelines for ethical review of epidemiological studies. *Council for International Organizations of Medical Sciences*, 1991.

[4] A. Alkhateeb, T. Takahashi, S. Mandil, and Y. Sekita. The changing role of health care IC card systems. *Proc. Computer Methods and Programs in Biomedicine*, 60(2):83–92, 1999.

[5] G. Antoniol and P. Tonella. EEG data compression techniques. *IEEE Trans. Biomedical Engineering*, 44(2):105 – 114, 1997.

[6] U. R. Archarya, D. Anand, P. S. Bhat, and U. C. Niranjan. Compact storage of medical images with patient information. *IEEE Trans. Information Technology in Biomedicine*, 5:320–323, 2001.

[7] B. S. Atal. The history of linear prediction. *IEEE Signal Processing Magazine*, 23(2):154–161, 2006.

[8] C. Bandt. Ordinal time series analysis. *J. Ecological Modelling*, 182:229–238, 2005.

[9] F. Bao, R. H. Deng, B. C. Ooi, and Y. Yang. Tailored reversible watermarking schemes for authentication of electronic clinical atlas. *IEEE Trans. Information Technology in Biomedicine*, 9(4):554–563, 2005.

[10] P. Bas and J. Hurri. Vulnerability of DM watermarking of non-IID host signals to attacks utilising the statistics of independent components. *IEEE Trans. Information Forensics and Security*, 153(3):127 –139, 2006.

[11] BBC. http://news.bbc.co.uk/1/hi/uk/7158019.stm.

[12] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3&4):313–336, 1996.

[13] C. M. Bishop. In *Neural networks for pattern recognition*. Oxford University Press, 1995.

[14] T. Blanchett, G. C. Kember, and G. A. Fenton. KLT-based quality controlled compression of single-lead ECG. *IEEE Trans. Biomedical Engineering*, 45(7):942–945, 1998.

[15] G. Boato, F. G. B. De Natale, and C. Fontanari. Digital image tracing by sequential multiple watermarking. *[IEEE] Transactions on Multimedia*, 9(4):677–686, 2007.

[16] S. Bounkong, B. Toch, D. Saad, and D. Lowe. ICA for watermarking. *J. Machine Learning Research*, 4(7-8):1471–1498, 2004.

[17] CanadaHealthInfoway. Annual report, 2005-2006. www.infoway-inforoute.ca/Admin/Upload/Dev/Document/Annual%20Report%2005-06%20EN.pdf.

[18] F. Cayre, C. Fontaine, and T. Furon. A theoritical study of watermarking security. *Proc. Information Theory (ISIT)*, pages 1868 –1872, 2005.

[19] H. Chao, C. Hsu, and S. Miaou. A data hiding technique with authentication, integration, and confidentiality for electronic patient records. *IEEE Trans. Information Technology in Biomedicine*, 6(1):46–53, 2002.

[20] B. Chen and G. W. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Information Theory*, 47(4):1423–1443, 2001.

[21] Y. Chen. Taiwan's health IC smart card security and privacy policy, 2003. Translated from www.martsoft.com/reference/healthcare/tw_nhi_2003Sep.pdf.

[22] C. S. Chuang. Human rights concern in an information society-thoughts on personal data protection in Taiwan. *Proc. The World Summit on the Information Society, Asian Regional Conference*, pages 277–315, 2003.

[23] I. J. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. In *Digital Watermarking and Steganography*. Morgan Kaufmann Publishers, 2007.

[24] I. J. Cox, M. L. Miller, and J. A. Bloom. In *Digital Watermarking*. Morgan Kaufmann Publishers, 2002.

[25] I.J. Cox, J. Killian, T. Leighton, and T. Shamoon. A secure, robust watermark for multimedia. *Proc. Workshop on Information Hiding*, pages 175–190, 1996.

[26] S. Craver, N. Memon, B. Yeo, and M.M. Yeung. Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications. *IEEE J. Selected Areas in Communications*, 16(4):573–586, 1998.

[27] S. Dandapat, O. Chutatape, and S. M. Krishnan. Perceptual model based data embedding in medical images. *Proc. Int. Conf. Image Processing (ICIP)*, 4:2315–2318, 2004.

[28] J.F. Delaigle, C. Devleeschouwer, B. Macq, and I. Langendijk. Human visual system features enabling watermarking. *Proc. IEEE Int. Conf. Multimedia and Expo*, 2:489 – 492, 2002.

[29] DL7. http://www.biopattern.org/Main/wfShowPubDocs.aspx.

[30] M. Eichelberg, T. Aden, J. Reismeier, A. Dogac, and G. B. Laleci. A survey and analysis of electronic healthcare record standards. *ACM J. Computing Surveys*, 37(4):277–315, 2005.

[31] C. Fontaine and P. Bas. Kerckhoffs-based embedding security classes for woa data hiding. *IEEE Trans. Information Forensics and Security*, 3(1):1 –15, 2008.

[32] L. Fritsche, G. Lindemann, K. Schroeter, A. Schlaefer, and H. Neumayer. Implementation of a web-based electronic patient record for transplant recipients. *J Med Internet Res*, 1:e8, 1999. available http://www.jmir.org/1999/suppl1/e8.

[33] A. Giakoumaki, S. Pavlopoulos, and D. Koutsouris. Multiple image watermarking applied to health information management. *IEEE Trans. Information Technology in Biomedicine*, 10(4):722–732, 2006.

[34] D. Q. Goldin and P. C. Kanellakis. On similarity queries for time-series data: constraint specification and implementation. *First Int. Conf. Principles and Practice of Constraint Programming, Lecture Notes in Computer Science*, 976:137–153, 1995.

[35] F. J. González-serrano, H. Y. Molina-Bulla, and J. J. Murillo-Fuentes. Independent component analysis applied to digital image watermarking. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 3:1997 – 2000, 2001.

[36] D. M. Green and J. A. Swets. In *Signal detection theory and psychophysics*. John Wiley and Sons, 1966.

[37] C. Le Guillou, G. Coatrieux, J-. M. Cauvin, L. Lecornu, and Ch. Roux. Enhancing shared medical image functionalities with image knowledge digest and watermarking. *Proc. Int. Special Topic Conference on Information Technology Applications in Biomedicine (ITAB 2006)*, 2006. http://medlab.cs.uoi.gr/itab2006/proceedings/MedicalImaging.htm.

[38] M. Haas. Healthconnect. *Health Policy Monitor*, 2005. Available at http://www.hpm.org/survey/au/a5/2.

[39] S. Haykin. In *Adaptive Filter Theory*. Prentice Hall Information and System Sciences Series, 1996.

[40] M. Holliman, N. Memon, and M. Yeung. Watermark estimation through local pixel correlation. *Proc. SPIE Security and watermarking of multimedia content I*, 3675:134 – 146, 1999.

[41] M. Hsieh, D. Tseng, and Y. Huang. Hiding digital watermarks using multiresolution wavelet transform. *IEEE Trans. Industrial Electronics*, 48(5):875–882, 2001.

[42] A. Hyvarinen, J. Karhunen, and E. Oja. In *Independent Component Analysis*. Wiley-Interscience, 2001.

[43] IHE. http://www.ihe.net/Technical_Framework.

[44] S. Jain. Digital watermarking techniques: A case study in fingerprints & faces. *Proc. Indian Conf. Computer Vision, Graphics, and Image Processing*, pages 139–144, 2000.

[45] N.J. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of the human perception. *Proc. of the IEEE*, 81(4):1385–1422, 1993.

[46] C. Jin and L. Pan T. Su. Multiple digital watermarking scheme based on ICA. *Proc. IEEE 8th Int. Workshop on Image Analysis for Multimedia Interactive Services*, pages 70–73, 2007.

[47] N. F. Johnson, Z. Duric, and S. Jajodia. Information hiding, steganography and watermarking - attacks and countermeasures. In *Information Hiding*. Kluwer Academic Publishers, 2000.

[48] N. F. Johnson and S. Katzenbeisser. In *Information Hiding*. Artech House, 2000.

[49] A. Kalja, A. Reitsakas, and N. Saard. eGovernment in Estonia: Best practices. *IEEE J. Tech. Management: A Unifying Discipline for Melting The Boundaries*, pages 500 – 506, 2005.

[50] T. Kalker. Considerations on watermarking security. *Proc. Multimedia Signal Processing, IEEE Fourth Workshop on*, pages 201–206, 2001.

[51] S. Kirshner and B. Póczos. ICA and ISA using schweizer-wolff measure of dependence. *Proc. 25th Int. Conf. Machine Learning, ACM Int. Conf. Proc. Series*, 307:464–471, 2008.

[52] G. Kurtz. EMR confidentiality and information security. *J. Healthcare Information Management*, 17(3):41–48, 2003.

[53] S. Lin and D. J. Costello Jr. In *Error control coding: fundamentals and applications*. Prentice-Hall computer applications in electrical engineering series, 1983.

[54] B. Macq and F. Dewey. Trusted headers for medical images. *DFG VIII - D II Watermarking Workshop*, 1999.

[55] B. R. Matam and D. Lowe. Steganography, biopatterns and independent components. *Proc. 7th Int. Conf. Mathematics in Signal Processing*, pages 206–209, 2006.

[56] B. R. Matam and D. Lowe. Standardisation of medical formats and protection of patient privacy. *Proc. 4th Conf. on Medical Data Acquisition Standards and Protocols*, 2007.

[57] B. R. Matam and D. Lowe. Towards gaining patient confidence in, and acceptance of the electronic patient health record: A steganographic proposal. *Proc. 3rd Int. Conf. Computational Intelligence in Medicine and Healthcare (CIMED)*, 2007.

[58] B. R. Matam and D. Lowe. Exploiting sensitivity of nonorthogonal joint diagonalisation as a security mechanism in steganography. *Proc. 16th Int. Conf. Digital Signal Processing*, 2009.

[59] B. R. Matam and D. Lowe. Participatory ephr: A watermarking solution. *Proc. 10th Int. Work-Conf. Artificial Neural Networks*, 2009.

[60] B. R. Matam and D. Lowe. Watermarking: How secure is the dm-qim watermarking technique? *Proc. 16th Int. Conf. Digital Signal Processing*, 2009.

[61] D. McKay. In *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2002.

[62] R. Mehul and R. Priti. Discrete wavelet transform based multiple watermarking scheme. *Proc. IEEE Region 10 Technical Conference on Convergent Technologies for the Asia-Pacific, Bangalore, India*, 2003.

[63] N. Memon, X. Kong, and J. Cinkler. Context-based lossless and near-lossless compression of EEG signals. *IEEE Trans. Information Technology in Biomedicine*, 3(3):231 –238, 1999.

[64] M. Milanova, C. Ford, R. Kountchev, and R. Kountcheva. Digital watermarking for medical images. *Proc. Int. Conf. on Mathematics and Engineering Techniques in Medicine and Biological Scienes, (METMBS*, pages 509–520, 2003.

[65] M. L. Miller, G. J. Doerr, and I. J. Cox. Applied informed coding and embedding to design a robust high-capacity watermark. *IEEE Trans. Image Processing*, 13:792–807, 2004.

[66] F. Mintzer and G. Braudaway. If one watermark is good, are more better. *Proc. Int. Conf. Accoustics, Speech, and Signal Processing*, 4:2067–2069, 1999.

[67] P. Moulin and J.A. O'Sullivan. Information-theoretic analysis of information hiding. `http://www.ifp.uiuc.edu/ moulin/`, preprint Sept. 1999, revised, December 2001.

[68] I. T. Nabney. In *Netlab, algorithms for pattern recognition*. Springer, 2002.

[69] NationalHealthService. Connecting for health, a guide to npfit, 2005. http://www.connectingforhealth.nhs.uk/publications/brochures/npfit_brochure_apr_05_final.p

[70] J. Nayak, U. R. Archarya, P. S. Bhat, and U. C. Niranjan. Simultaneous storage of medical images in the spatial and frequency domain: a comparative study. *Biomedical Engineering Online*, 2004. available http://www.biomedical-engineering-online.com/content/3/1/17.

[71] R. T. Ogden. In *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser Boston, 1997.

[72] L. Ohno-Machado, P. S. P. Silveira, and S. Vinterbo. Protecting patient privacy by quantifiable control of disclosures in disseminated databases. *Int. J. Med Informatics*, 73:599–606, 2004.

[73] openEHR. Copyright openEHR Foundation 2001-2006 www.openEHR.org.

[74] D. B. Percival and A. T. Walden. In *Wavelet methods for time series analysis*. Cambridge Series in Statistical and Probabilistic MathematicsCambridge University Press, 2000.

[75] L. Pérez-Freire and F. Pérez-González. Exploiting security holes in lattice data hiding. *Proc. Information Hiding*, pages 159–173, 2007.

[76] L. Pérez-Freire, F. Pérez-González, T. Furon, and P. Comesaña. Security of lattice-based data hiding against the known message attack. *IEEE Trans. Information Forensics and Security*, 1(4):421 –439, 2006.

[77] F. A. P. Petitcolas. Watermarking schemes evaluation. *IEEE Signal Processing Magazine*, pages 58–64, 2000.

[78] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn. Information hiding - a survey. *Proc. of the IEEE*, 87(7):1062– 1078, 1999.

[79] B. M. Planitz and A. J. Maeder. Medical image watermarking: A study on image degradation. *Proc. Workshop on Digital Image Computing (WDIC 2005), Brisbane, Australia*, pages 3–8, 2005.

[80] B. M. Planitz and A. J. Maeder. A study of block-based medical image watermarking using a perceptual similarity metric. *Proc. Digital Image Computing: Techniques and Applications (DICTA 2005)*, pages 483–490, 2005.

[81] I. Popivanov and R. J. Miller. Similarity search over time-series data using wavelets. *Proc.* 18^{th} *Int. Conf. on Data Engineering*, pages 02–12, 2002.

[82] J. G. Proakis and D. G. Manolakis. In *Digital Signal Processing*. Prentice-Hall India, 1999.

[83] D. Rafiei and A. Mendelzon. Similarity-based queries for time series data. *Proc. SIGMOD Conference*, pages 13–25, 1997.

[84] N. Ramkumar and N. Memon. making a mark. *spie's oe magazine*, pages 20–23, 2003.

[85] O. Rienhoff, R. J. Rodrigues, U. Piccolo, A. Hernandez, and N. Oliveri. Integrated circuit health data cards (smart cards): A primer for health professionals. *Technology and Health Services Delivery, Health Services Organization Unit (THS/OS), Pan American Health Organization*, 2003.

[86] P. Russak. One card to rule them all - national identity card. *Proc. EUNIS 2005 conference -Leadership and Strategy in a Cyber-Infrastructure World*, 2005.

[87] H. T. Sencar and N. Memon. Combatting ambiguity attacks via selective detection of embedded watermarks. *IEEE Trans. Information Forensics and Security*, 2:664–682, 2007.

[88] N. P. Sheppard and P. Ogunbona R. Safavi-Naini. On multiple watermarking. *Proc. ACM Workshop on Multimedia and Security: new challenges*, pages 3–6, 2001.

[89] W. Stallings. In *Cryptography and Network Security Principles and Practice*. Pearson Education International, 2003.

[90] I. Stevenson and D. Ensor. In *Oracle Design: The Definitive Guide*. O'Reilly, 1997.

[91] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. S. Manjunath. Steganalysis of quantization index modulation data hiding. *Proc. Int. Conf. Image Processing*, 2:1165 – 1168, 2004.

[92] K. Sullivan, K. Solanki, B. S. Manjunath, U. Madhow, and S. Chandrasekaran. Determining achievable rates for secure, zero divergence, steganography. *Proc. IEEE Int. Conf. on Image Processing*, pages 121–124, 2006.

[93] M.D. Swanson, M. Kobayashi, and A.H. Tewfik. Multimedia data-embedding and watermarking technologies. *Proc. of the IEEE*, 86(6):1064–1087, 1998.

[94] P. Tao and A. M. Eskicioglu. A robust multiple watermarking scheme in the discrete wavelet transform domain. *Proc. SPIE Internet Multimedia Management Systems V*, 5601:133–144, 2004.

[95] D. Tian and M. Ha. Applications of wavelet transform in medical image processing. *Proc. Third International Conference on Machine Learning and Cybernetics*, 3:1816 – 1821, 2004.

[96] B. Toch and D. Lowe. Watermarking of medical signals. *Proc.* 2nd *Int. Conf. Computational Intelligence in Medicine and Healthcare*, pages 231–236, 2005.

[97] B. Toch, D. Lowe, and D. Saad. Watermarking of audio signals using independent component analysis. 3rd *International Conference on WEB Delivering of Music (WEDELMUSIC'03)*, pages 71–74, 2003.

[98] Y. Wang, J. F. Doherty, and R. E. Van Dyck. A wavelet-based watermarking algorithm for ownership verification of digital images. *IEEE Trans. Image Processing*, 11(2):77 – 88, 2002.

[99] P. H. W. Wong and O. C. Au A. Chang. A sequential multiple watermarks embedding technique. *Proc. Int. Conf. on Accoustics, Speech and Signal Processing*, 5:393–396, 2004.

[100] C. Woo, J. Du, and B. Pham. Multiple watermark method for privacy control and tamper detection in medical images. *Proc. APRS Workshop on Digital Image Computing (WDIC 2005)*, 2005.

[101] W. L. Woon and D. Lowe. Nonlinear signal processing for noise reduction of unaveraged single channel MEG data. *Proc. Int. Conf. Artificial Neural Networks*, pages 650 –657, 2001.

[102] J. M. Zain and A. M. Fauzi. Medical image watermarking with tamper detection and recovery. *Proc. IEEE Conf. on Engineering in Medicine and Biological Sciences*, pages 3270–3273, 2006.