

## ABSTRACT

Title of dissertation:       TIME AND LOCATION FORENSICS  
FOR MULTIMEDIA

Ravi Garg, Doctor of Philosophy, 2013

Dissertation directed by:  Professor Min Wu

Department of Electrical and Computer Engineering

In the modern era, a vast quantities of digital information is available in the form of audio, image, video, and other sensor recordings. These recordings may contain metadata describing important information such as the time and the location of recording. As the stored information can be easily modified using readily available digital editing software, determining the authenticity of a recording has utmost importance, especially for critical applications such as law enforcement, journalism, and national and business intelligence.

In this dissertation, we study novel environmental signatures induced by power networks, which are known as Electrical Network Frequency (ENF) signals and become embedded in multimedia data at the time of recording. ENF fluctuates slightly over time from its nominal value of 50 Hz/60 Hz. The major trend of fluctuations in the ENF remains consistent across the entire power grid, including when measured at physically distant geographical locations. We investigate the use of ENF signals for a variety of applications such as estimation/verification of time and location of a recording's creation, and develop a theoretical foundation to support ENF based forensic analysis.

In the first part of the dissertation, the presence of ENF signals in visual recordings captured in electric powered lighting environments is demonstrated. The source of ENF signals in visual recordings is shown to be the invisible flickering

of indoor lighting sources such as fluorescent and incandescent lamps. The techniques to extract ENF signals from recordings demonstrate that a high correlation is observed between the ENF fluctuations obtained from indoor lighting and that from the power mains supply recorded at the same time. Applications of the ENF signal analysis to tampering detection of surveillance video recordings, and forensic binding of the audio and visual track of a video are also discussed.

In the following part, an analytical model is developed to gain an understanding of the behavior of ENF signals. It is demonstrated that ENF signals can be modeled using a time-varying autoregressive process. The performance of the proposed model is evaluated for a timestamp verification application. Based on this model, an improved algorithm for ENF matching between a reference signal and a query signal is provided. It is shown that the proposed approach provides an improved matching performance as compared to the case when matching is performed directly on ENF signals. Another application of the proposed model in learning the power grid characteristics is also explicated. These characteristics are learnt by using the modeling parameters as features to train a classifier to determine the creation location of a recording among candidate grid-regions.

The last part of the dissertation demonstrates that differences exist between ENF signals recorded in the same grid-region at the same time. These differences can be extracted using a suitable filter mechanism and follow a relationship with the distance between different locations. Based on this observation, two localization protocols are developed to identify the location of a recording within the same grid-region, using ENF signals captured at anchor locations. Localization accuracy of the proposed protocols are then compared. Challenges in using the proposed technique to estimate the creation location of multimedia recordings within the same grid, along with efficient and resilient trilateration strategies in the presence of outliers and malicious anchors, are also discussed.

TIME AND LOCATION FORENSICS  
FOR MULTIMEDIA

by

Ravi Garg

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2013

Advisory Committee:  
Professor Min Wu, Chair/Advisor  
Professor K. J. Ray Liu  
Professor Gang Qu  
Dr. Avinash L. Varna  
Professor Lawrence C. Washington

© Copyright by  
Ravi Garg  
2013

*To my parents.*

## ACKNOWLEDGEMENTS

I owe my sincere gratitude to all the people who made this dissertation see a light of the day. All these people made my graduate study an experience that I will cherish forever.

First and foremost I'd like to express my deepest gratitude to my advisor, Professor Min Wu, for giving me an opportunity to work on an exciting array of challenging and extremely interesting research problems. Her constant guidance and support helped me in making a steady progress to achieve my research goals. She has always made herself available for help and advice, not only on the research problems but also to shape my career plans. There has never been an occasion when I have knocked on her door, and she has not lent a helping hand. Her invaluable advice has helped in improving my writing and presentation skills. I will definitely be using throughout my life the lessons that I have learnt from her about dedication towards work, time management, and to maintain a positive attitude in life.

I would like to thank Prof. K. J. Ray Liu for the outstanding teaching in his courses, which shaped up my decision to conduct research in the signal processing domain. I would also like to thank him for his unwavering support during my graduate study and serve on my dissertation committee. I would like to thank Prof. Gang Qu and Prof. Lawrence Washington to serve on my dissertation committee and provide their invaluable comments. I would like to thank Dr. Avinash Varna, with whom I spent three wonderful years of my graduate studies (as a MASTer), for serving on my dissertation committee. His mentorship has helped me in conducting my research and shaping up this dissertation. I also appreciate the mentorship and financial support offered to me by the Clark School of Engineering's Future Faculty Program, the Graduate School's Fellowship, and the Bulgaria Summer Research Fellowship at Maryland.

I am privileged to have worked in the company of some of the outstanding

colleagues at Maryland. I want to thank my fellow MAST members Dr. Wenjun Lu, Dr. Wei-Hong Chuang, David Hou, Hui Su, Adi-Hajj Ahmad, Chau-Wai Wong, Abbas Kazemipour, with whom I had many interesting discussions about many diverse topics. Thanks to all the friends I made at Maryland. Intense squash sessions with Dr. Kapil Anand, Shalabh Jain, Vaibhav Singh, Krishna Chaitanya, Sumit Sekhar were always refreshing. Discussions with Dr. Amardip Ghosh and Udayan Khurana were very enlightening. Frequent DC late night outings with Dr. Amit Goyal and Amit Mathur were very much enjoyable. Thanks to all my roommates, Dr. Pavan Soma, Dr. Bhargava Ravoori, Rohan Kapoor, Ishwinder Singh, Akhil Anand, Mitabh Patel for cooking great food. For my whole life, I would cherish the memories of the wonderful time that I spent with my friends at Maryland.

Last but not the least, I owe my greatest thank to the people who love me the most. I owe my sincerest gratitude to my parents for their constant support and encouragement to pursue my ambitions throughout my life. I would like to thank my sisters, my brother, and other member of my family for their continuing support. I dedicate this dissertation to them.

# Table of Contents

List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Motivation . . . . .	1
1.2 Background . . . . .	3
1.3 Main Contribution and Dissertation Organization . . . . .	6
2 Electric Network Frequency based Timestamping in Optical Sensor Recordings	11
2.1 Chapter Introduction . . . . .	11
2.2 ENF Signal Extraction in Indoor Lightings . . . . .	13
2.2.1 ENF Extraction using Optical Sensors . . . . .	15
2.2.2 Instantaneous Frequency Estimation . . . . .	18
2.2.2.1 Non-Parametric Methods . . . . .	18
2.2.2.2 Parametric Methods . . . . .	20
2.2.3 Correlation Between Sensor Signal and Power Signal . . . . .	22
2.2.4 Experimental Results and Discussions . . . . .	24
2.2.4.1 Experiment 1 . . . . .	24
2.2.4.2 Experiment 2 . . . . .	28
2.2.4.3 Experiment 3 . . . . .	29
2.3 Chapter Summary . . . . .	31
3 Video Forensics using ENF Signals	32
3.1 Chapter Introduction . . . . .	32
3.2 ENF Extraction from Video . . . . .	34
3.2.1 Aliasing Analysis . . . . .	35
3.2.2 Experimental Setup for Video Recordings . . . . .	38
3.2.2.1 White Wall Video . . . . .	39
3.2.2.2 Surveillance Video . . . . .	39
3.3 Applications of ENF Signal Analysis to Video Forensics . . . . .	40



3.3.1	Estimating or Verifying Time-of-Recording . . . . .	41
3.3.1.1	Recordings from China . . . . .	42
3.3.1.2	Recordings from India . . . . .	45
3.3.1.3	Recordings from USA . . . . .	48
3.3.2	Tampering Detection . . . . .	50
3.3.3	Audio-Visual Authentication and Synchronization . . . . .	53
3.4	Discussions and Extensions . . . . .	54
3.4.1	Effect of Compression . . . . .	54
3.4.2	ENF Extraction from CMOS Imaging Cameras . . . . .	55
3.5	Chapter Summary . . . . .	59
4	Statistical Modeling and Analysis of ENF Signals . . . . .	61
4.1	Chapter Introduction . . . . .	61
4.2	Autoregressive (AR) Model for ENF . . . . .	63
4.2.1	Statistics of ENF signals . . . . .	64
4.2.2	Box-Jenkins Test for Model Validation . . . . .	67
4.3	Timestamp Verification as Hypothesis Testing . . . . .	75
4.3.1	Matching using ENF sequences . . . . .	76
4.3.2	Results & Discussions . . . . .	78
4.4	Improved Method for Matching ENF Signals . . . . .	83
4.4.1	Results on Audio . . . . .	85
4.4.2	Results on Video . . . . .	89
4.5	Chapter Summary . . . . .	91
5	AR Model Parameters for Grid of Recording Classification . . . . .	95
5.1	Chapter Introduction . . . . .	95
5.2	ENF Variations Across Different Grids . . . . .	97
5.3	AR Modeling Parameters as Features for Location Classification . . . . .	100
5.3.1	Feature Description . . . . .	100
5.3.2	Experimental Setup . . . . .	102
5.3.3	Results and Discussions . . . . .	104
5.4	Noise Adaptation using Multi-Conditional Learning . . . . .	109
5.4.1	System Model . . . . .	110
5.4.2	Results and Discussions . . . . .	117
5.5	Chapter Summary . . . . .	122
6	Intra-grid Location Estimation using ENF Signals . . . . .	125
6.1	Chapter Introduction . . . . .	125
6.2	Propagation Mechanism of ENF Signal . . . . .	126
6.3	Location Dependence of ENF Signals . . . . .	128
6.3.1	Signal Processing Mechanism to Extract ENF Variations . . . . .	129
6.3.2	Case Study 1: 3-Location Data on the US East Coast . . . . .	130
6.3.3	Case Study 2: 5-Location Data on the US East Coast . . . . .	134
6.4	Half-Plane Intersection for Localization . . . . .	137
6.5	Correlation Quantization for Localization . . . . .	143

6.6	Sensitivity Analysis . . . . .	148
6.7	Chapter Summary . . . . .	151
7	A Gradient Descent Approach to Secure Localization . . . . .	153
7.1	Chapter Introduction . . . . .	153
7.1.1	Prior Work . . . . .	155
7.2	Secure Localization in Wireless Sensor Networks . . . . .	158
7.2.1	Problem Formulation . . . . .	158
7.2.1.1	Non-coordinated Attacks . . . . .	160
7.2.1.2	Coordinated Attacks . . . . .	160
7.2.2	Proposed Method for Secure Localization . . . . .	161
7.2.3	Comparison of Computational Complexity . . . . .	165
7.2.4	Simulation Results . . . . .	170
7.2.4.1	Non-coordinated attacks . . . . .	171
7.2.4.2	Coordinated attacks . . . . .	173
7.2.5	Discussions . . . . .	175
7.3	Gradient Descent Approach Applied to TDoA Measurements . . . . .	179
7.3.1	Secure Localization Problem for TDoA . . . . .	182
7.4	Chapter Summery . . . . .	186
8	Conclusions & Future Perspectives . . . . .	188
	Bibliography . . . . .	193

## List of Tables

3.1	Aliased frequency for different combinations of power mains frequencies and video camera frame-rates . . . . .	38
4.1	Hypothesis model of Box-Jenkins methodology. . . . .	68
5.1	Number of training and testing examples in the dataset . . . . .	105
5.2	Classification accuracy for power-ENF training . . . . .	107
5.3	Classification accuracy on power-ENF training for all the methods . .	120
7.1	Comparison of run time complexity of different localization algorithms	165

## List of Figures

2.1	Light sensing circuit used to verify presence of ENF in indoor lightings	17
2.2	Block diagram for estimating creation time . . . . .	23
2.3	ENF fluctuations of signals captured in Experiment 1 . . . . .	24
2.4	ENF fluctuations using different frequency estimation methods . . . . .	26
2.5	NCC coefficient, $\rho(k)$ , for experiment 1 . . . . .	27
2.6	ENF for optical sensor signal in Experiment 2 . . . . .	29
2.7	NCC coefficient, $\rho(k)$ , for experiment 3 . . . . .	30
3.1	Illustration of the aliasing effect in video capturing of ENF signals . . . . .	36
3.2	ENF fluctuations for whitewall video experiment in China . . . . .	41
3.3	$\rho(k)$ using different frequency estimation methods on China data . . . . .	43
3.4	ENF fluctuations for whitewall video experiment in China . . . . .	45
3.5	ENF fluctuations measured for the whitewall video experiment in India . . . . .	47
3.6	Average Peak NCC value for matching and non-matching case . . . . .	48
3.7	ENF fluctuations measured for surveillance video experiment in India . . . . .	49
3.8	ENF fluctuations measured for the whitewall video experiments in US . . . . .	50
3.9	Video tampering detection using ENF signals . . . . .	52
3.10	ENF matching between audio and video tracks of a video recording . . . . .	54
3.11	Peak NCC value for different compression rates . . . . .	55
3.12	Rolling shutter sampling mechanism in CMOS cameras . . . . .	57
3.13	Spectrograms of a video signal recorded using a CMOS camera . . . . .	57
3.14	Average peak NCC for the videos recorded using CMOS image sensor . . . . .	59
4.1	The mean and the autocorrelation function of an ENF signal recording . . . . .	66
4.2	pdf of $f(n)$ for a 25-hours long power-ENF recording . . . . .	66
4.3	Plot of autocorrelation function. . . . .	68
4.4	sample PACF $\alpha(k)$ . . . . .	70
4.5	Distribution of partial autocorrelation function $\alpha(k)$ . . . . .	70
4.6	Distribution of estimated AR(2) parameters . . . . .	71
4.7	Mean and variance of $Q(m)$ for AR(2) model of ENF signals . . . . .	73
4.8	Mean and variance of $Q(m)$ for different order of AR model of ENF signals . . . . .	74

4.9	AIC(k) and BIC(k) for the residue process . . . . .	75
4.10	Receiver Operating Characteristics (ROC) of the hypothesis detection framework for timestamp verification using the proposed model . . .	79
4.11	Comparison of the ROCs of the hypothesis detection framework for $T_{frame} = 8$ seconds and $T_{frame} = 16$ seconds . . . . .	80
4.12	Pdf of $S$ for the audio-power data for $T_{frame} = 16$ seconds under $H_0$ and $H_1$ hypotheses for different query durations . . . . .	81
4.13	Comparison of the ROCs of the hypothesis detection framework for timestamp verification under the proposed model and the empirical data for different query duration . . . . .	82
4.14	Receiver Operating Characteristics(ROC) of the proposed AR decorrelation based hypothesis detection framework for timestamp verification using the model . . . . .	86
4.15	Pdf of $S'$ for the audio-power data $T_{frame} = 16$ seconds under $H_0$ and $H_1$ hypotheses for different query durations . . . . .	87
4.16	ROC characteristics of the correlation detector for ENF matching vs. innovation matching at two query segment length for $T_{frame} = 16$ seconds . . . . .	89
4.17	ROC for timestamp verification when using ENF signal and innovation sequences for matching . . . . .	90
5.1	Sample ENF plots for European and Asian grids . . . . .	98
5.2	Sample ENF plots for continental North American grids . . . . .	99
5.3	AR parameters for location-of-recording classification . . . . .	103
5.4	Classification accuracy for RBF kernel using 10-fold cross validation .	105
5.5	Classification accuracy on power testing data . . . . .	108
5.6	Classification accuracy for different training and testing conditions . .	109
5.7	Block diagram of multi-conditional learning for noise adaptation . . .	111
5.8	Parzen-window density estimation for different noise conditions . . . .	116
5.9	Classification accuracy using the GMM based Bayesian classifier . . . .	118
5.10	Classification accuracy for different test conditions using the GMM based Bayesian classifier . . . . .	119
5.11	Classification accuracy comparison for different Bayesian approaches .	120
5.12	ROC of the proposed location verification detector . . . . .	122
6.1	Sample ENF signals from three location recordings in US east . . . .	129
6.2	Signal processing mechanism to extract intra-grid ENF Signatures . .	130
6.3	Correlation coefficient between processed ENF signals for 3-location data in US east coast . . . . .	131
6.4	Three locations shown on a map for Case Study 1 . . . . .	132
6.5	Mean error in distance estimation between Princeton and Atlanta data	133
6.6	Five locations shown on a map for Case Study 2 . . . . .	135
6.7	Correlation coefficient between the processed ENF signals for 5-location data on US east . . . . .	136
6.8	Grid density of the US east interconnection grid . . . . .	137

6.9	Localization example for Princeton using half-plane intersection . . .	140
6.10	$p_{loc}$ and $a_{loc}$ for 5-location US east data using the half-plane intersection method . . . . .	142
6.11	Relationship between $\rho_{i,j}$ and $d_{i,j}$ for 5-location data . . . . .	142
6.12	Localization example for Princeton using the correlation quantization method. . . . .	145
6.13	Localization example for Princeton using three localization protocols	147
6.14	$p_{loc}$ and $a_{loc}$ using different localization methods . . . . .	149
6.15	$p_{loc}$ under different noisy conditions using half-plane intersection method	150
7.1	Force Vector representation of terms contributing to the gradient . . .	163
7.2	Flow chart for localization algorithm . . . . .	167
7.3	Comparison of run-time for different localization schemes for a fixed localization error . . . . .	169
7.4	Comparison of localization schemes for non-coordinated attacks . . .	171
7.5	Probability of converging to the correct estimate . . . . .	172
7.6	Performance of the secure localization schemes under coordinated attacks by 30% of the nodes . . . . .	174
7.7	Geometry for the bound on the maximum number of colluding nodes	177
7.8	Diagram representing the basic TDoA protocol . . . . .	181
7.9	Intersection of three hyperbolas gives the location of a node when TDoA is used . . . . .	183
7.10	Localization accuracy for coordinated attacks using TDoA measurements . . . . .	186
7.11	Probability of converging to the correct estimate for TDoA . . . . .	187

# Chapter 1

## Introduction

### 1.1 Motivation

In the modern era, a huge amount of digital information is available in the form of audio, image, video, and other sensor recordings. These recordings may contain metadata describing important information such as the time and the place of recording. However, digital tools can be used to modify the stored information. For example, digital editing softwares can be used to cut a clip from an original audio or to insert a foreign clip into it, or to manipulate the metadata field to alter the recording date and time. Similar changes can also occur with video surveillance and other types of recordings. These modifications can result in serious consequences when multimedia recordings are used for law enforcement and journalism cases. Multimedia forensics is gaining importance in the information era in light of these capabilities.

The questions related to the time and place of recordings, and their authen-

ticity also have grave importance for counter-terrorism operations. In recent years, terrorists like Osama bin Laden and other have released many propaganda video and audio recordings threatening attacks on the civilians and governmental infrastructure. To detain the perpetrators of such crimes, investigation agencies seek answers to questions such as: When was the given recording created? Where was it created? Is the recording authentic? To the best of our knowledge, no satisfactory technology exists that may be applied generally and potentially answer such questions. This is a major impediment toward counter-terrorism operations.

Several classes of signal processing techniques have been discussed in literature to verify the integrity of multimedia recordings. For example, watermarking is a proactive measure to achieve data security and involves embedding the hosting signal with a signal known to the system designer [1]. In order to determine the authenticity of a recording in question, a forensics examiner can extract the watermark from the recording and compare it with the database. One of the limitations of digital watermarking techniques is that the watermark signal is embedded at the time of data creation or data transmission, and it requires dedicated software/hardware and protocol. In the absence of proactive watermarking techniques during the initial data acquisition, an adversary can easily modify the information in a given recording to his/her advantage.

Another class of signal processing techniques developed for multimedia forensics relates to non-intrusive forensics that traces the origin and the processing history of multimedia content to its creation process [2]. For example, component forensics aims to identify the algorithms and parameters employed in the various components



of the device used in capturing the data [3]. These techniques work by identifying the traces left in multimedia data when it goes through various processing blocks inside the recording device. Such techniques can address a number of forensics issues, such as in discovering device-technology infringement, protecting intellectual property rights, and identifying acquisition devices. This class of techniques offers a major and more recent alternative to watermarking because they do not require the active embedding of a known signal in the recording content, but alone they cannot address the questions related to the time and location of a multimedia recording. These limitations in the existing schemes motivates us to investigate alternative technologies to ensure information authenticity.

## 1.2 Background

The main challenges in multimedia forensics have been to determine the time and location of a given recording's creation. Forensic investigators seeking information about creation time and creation location mainly rely on the metadata, which may contain such information. Creation time can be stored with the metadata using built-in clocks in most modern media recording devices such as audio recorders, camcorders, etc. Location-of-recording can also be stored automatically with recording in devices equipped with a global positioning system (GPS), available in an increasing number of modern consumer cameras. However, with an advancement in digital tools, it is simple to modify this information stored in the metadata. Detecting such forgery has been a challenging task for forensic examiners. In this

thesis, we investigate novel solutions to these multimedia authentication and forensic problems by exploring the presence of an unique time and location dependent environmental signature, which is embedded in multimedia recordings at the time of their creation. These signatures arise from ubiquitous power networks, and are referred to as Electrical Network Frequency (ENF) signals.

ENF is the supply frequency of power distribution networks in a power grid. The nominal value of the ENF is 60 Hz in the United States and most parts of the Americas; and 50 Hz in most areas of Europe and Asia. Japan uses dual standards of 50 Hz and 60 Hz; Taiwan uses 60 Hz in its power distribution networks. An important property of the ENF comes from its fluctuations from the nominal value due to dynamically varying loads on the power grid [4]. To ensure the stability of a power network and proper functioning of many electric equipments, the grid should be tightly controlled to keep the fluctuations close to the nominal value. For example, in continental Europe, these fluctuations make the ENF a continuous random variable with typical values between 49.90 Hz and 50.10 Hz [4], and in the United States, ENF typically varies between 59.90 Hz and 60.10 Hz [5]. The main trends of fluctuations from the nominal value are consistent across the power grid's geographical region, and it has been shown that frequency fluctuations measured at a given time at locations physically far apart but connected to the same power grid have similar values [4] [5]. Similar variations in the ENF across the entire interconnected power grid stem from the load control mechanism used to stabilize the power grid [6]. Therefore, the ENF signal recorded at any location in a power grid at a given time can generally serve as a representative ENF signal in the geographical

area covered by the power grid at that time.

ENF signal traces are present in various types of multimedia recordings. Digital audio devices, such as microphone-based voice recorders that are plugged into the power mains or located near power sources or power transmission lines, often pick up the ENF signal because of interference from electromagnetic fields generated from power sources, acoustic hum, and mechanical vibrations produced by surrounding electric devices [4]. The nature of fluctuations embedded in digital recordings is similar to the fluctuations in the power signal captured directly from the power mains at the corresponding time. In recent years, this property of ENF signals has been explored to authenticate digital audio recordings [4] [5]. These works demonstrate that audio recordings capture ENF signal that exhibits a high correlation with the ENF signal recorded from the power mains supply at the corresponding time.

Complementary to the active recording of ENF signals, the phase continuity of ENF signals has been used to detect and locate segments of tampering in a given audio without resorting to the reference ENF signal. A sudden change in the ENF signal's phase extracted from audio indicates that the original recording has probably been modified [7]. This technique was also used to determine the authenticity of the White House tapes recorded during the Watergate scandal in 1970s [8]. The investigations discovered that the famous eighteen and one-half minute section of the recording with buzz sounds does not contain any tampering evidences with the tapes. This method does not require reference ENF to detect tampering, though determining the time-of-recording and the time of any inserted segments would still require matching against a database of reference ENF signals.

## 1.3 Main Contribution and Dissertation Organization

Although most of the prior work of applying ENF signals for forensics is limited to audio timestamping, a systematic statistical study of ENF signals has not been conducted in the audio forensics literature. Also, no examination of the presence of ENF signals in visual recordings has occurred. In this dissertation, we explore several uncharted aspects of ENF signal analysis, which enables us to answer the forensics questions related to the creation time and location of a recording. The main contribution of this dissertation are as follows:

- This dissertation presents the first study to demonstrate the presence of ENF signals in visual recordings. As visual recordings gain importance in many security and law enforcement applications, it becomes increasingly necessary to develop techniques to authenticate these recordings from multiple aspects of determining the time and location of the recording. We demonstrate the presence of ENF signal in videos recorded using consumer-end cameras, and develop signal processing techniques to extract ENF signals from such videos for timestamping, location-of-recording estimation, tampering detection, and audio-visual binding of a video.
- We conduct a statistical study of ENF signals with applications to multimedia timestamp verification. We also propose an analytical model for ENF signals based on an autoregressive (AR) process and demonstrate that ENF signals

fit closely into the proposed model. We use the proposed model to devise an improved method to match two ENF signals.

- In this dissertation, we also explore the location dependent properties of ENF signals. We demonstrate that the nature of ENF fluctuations varies across different grid regions. Based on this observation, we devise a location classification scheme that estimates the grid-region of recording of ENF influenced media recordings, using AR model parameters as features. We also explore the “microscopic” properties of ENF signals, which can be used to differentiate recordings conducted at the same time at different locations across the same power grid region. Using these properties, we devise localization protocols to estimate the location of recording within the same grid-region. For a complete exploration, we propose a localization scheme capable of providing excellent localization accuracy and computational efficiency even in the presence of adversaries injecting malicious location information.

The dissertation is organized as follows. In Chapter 2, we explore the source of ENF signals in visual recordings using experiments conducted on photo-diodes. We hypothesize that ENF signals are embedded in visual recordings from electric powered light sources such as fluorescent and incandescent lights. We devise experiments to verify our hypothesis, and propose methods to extract ENF signal from such recordings. We demonstrate that ENF signals can be used to timestamp optical sensor recording conducted under electric powered illuminated environment.

In Chapter 3, we explore the presence of ENF signals in video recordings. We demonstrate that ENF signals are embedded in the visual track of video recordings conducted in indoor environments illuminated by electric powered fluorescent lights. We discuss the challenges in extracting ENF signals from such recordings due to a low frame-rate of videos. We design a signal processing mechanism to extract ENF signals from video recordings and use them for various multimedia forensics applications such as timestamping, forgery detection of clip insertion and deletion, and audio-visual binding for videos.

In Chapter 4, we conduct a statistical study of ENF signals and propose a piecewise stationary autoregressive (AR) model to represent ENF signals. We validate the proposed model using the Box-Jenkins methodology, which is widely used in the literature of statistical time-series analysis. We use the proposed model to predict the performance of an ENF-based timestamping application under a binary hypothesis detection framework, and we demonstrate that the proposed model can predict the performance obtained on audio-power ENF database for a given signal-to-noise ratio (SNR) and query length. Using the proposed AR model, we also develop a new method to measure the similarity between two ENF signals. We demonstrate the effectiveness of the proposed method in terms of a performance improvement for timestamp verification application.

In Chapter 5, we explore another application of the AR model for ENF signals to learn the characteristics of different power grids. We demonstrate that ENF signal characteristics varies from one power grid to another. Based on this observation, we build a support vector machine based classifier to estimate the grid-region of

recording by using AR modeling parameters as features. We also evaluate the effect of noisy ENF signals on location classification accuracy and propose a Bayesian approach to noise adaptation, which provides a superior classification performance compared to when noise adaptation is not performed.

In Chapter 6, we study applications of ENF signal analysis for location-of-recording estimation within the same grid-region. We show that differences exist between the ENF signals captured directly from power mains at several locations of the same grid. We design a mechanism to extract location specific signatures from these recordings and demonstrate that suitable localization methods can be used to estimate the location-of-recording to a certain precision. We also discuss the challenges in using the proposed methodology to estimate the location-of-recording in multimedia recordings due to the noisy ENF signal embedding in such recordings, then we present a quantitative criterion to measure precision and resolution to improve in future work.

In Chapter 7, we propose a localization scheme that applies to ENF based trilateration and general sensor localization problems. The general problem assumes that the distances between the anchors and the sensor in question can be estimated. Some of these measurements may not be accurate given the presence of noise. Generally speaking, it is possible to have adversaries who attempt to inject malicious location information into the system. We develop an efficient method to localization under such settings and demonstrate that the proposed method can outperform the existing localization schemes in terms of computational requirements at a similar or better localization accuracy.

In Chapter 8, we conclude this dissertation and outline research issues for future explorations.



# Chapter 2

## Electric Network Frequency based Timestamping in Optical Sensor Recordings

### 2.1 Chapter Introduction

Electrical Network Frequency (ENF) is the supply frequency of power distribution networks in a power grid. ENF signal fluctuates from its nominal value of 50 Hz or 60 Hz due to dynamically varying loads on the power grid [4]. As ENF signals are also captured in audio recordings at the time of its creation, it has been used in a variety of forensics application such as timestamping and tampering detection in recent years [4] [5] [7]. For example, ENF signal from audio recordings can be estimated using signal processing mechanisms and compared with ground truth power mains signal to estimate the creation time of the given audio. The recorded signal is

filtered using a narrow bandpass filter with passband centered at the nominal ENF of interest. Spectrogram based frequency estimation algorithms are applied to extract the ENF fluctuations, which are compared with the ENF fluctuations from the power grid stored in a database to determine the time-of-recording or authenticity of audio. The ENF database may be obtained from the power utilities companies as they often keep track of it, or can be proactively recorded using simple hardware utilizing a step-down transformer and a voltage divider. ENF database can also be synthesized from multimedia recordings containing the ENF signal at a known recording time.

The next natural research question to address in the direction of using ENF fluctuations in multimedia forensics is: Are ENF signals present and measurable in visual sensing data? And what is the potential source of ENF in such measurements? CCD and CMOS imaging sensors used in video cameras measure the intensity of photons falling on the sensor array. These intensities are converted to measurable quantities, such as electric current or further into digital image sequences in video recordings. The intensity of light from light sources, such as fluorescent or incandescent bulbs, varies according to the frequency of the electric current, and may induce the ENF signal in recordings made using light sensors and visual sensors.

In this chapter, we explore the presence of ENF signals in optical sensor recordings in indoor lightings by conducting experiments on light capturing devices [9]. We build a light sensing device which produces current when light falls on it. The light sensing device is capable of capturing the light in the wavelength range of the common electricity powered lighting environments such as fluorescent lights and in-

candescent lights. We develop a method to detect the presence of the ENF signal from optical recordings conducted in indoor lightings and validate the timestamping property of the ENF signal for such recordings. We investigate the use of spectrum based frequency estimation methods and such subspace based frequency estimation algorithms as the MUSIC and the ESPRIT to estimate the ENF signal with high temporal resolution. Such a high temporal resolution of ENF estimation is beneficial for some emerging applications using ENF signal analysis. For example, it will be demonstrated in Chapter 6 that ENF signals contain location specific signatures even when they are recorded within the same grid; These signatures can be extracted using high temporal resolution frequency estimators to determine the location of a recording within the same grid [10] [11].

## 2.2 ENF Signal Extraction in Indoor Lightings

ENF signal is captured in the audio recordings because of interferences from electromagnetic field generated from nearby power sources and power transmission lines or the acoustic hum and mechanical vibrations produced by electric powered devices. Such observations raise the question: can ENF signal be captured in the visual recordings? To answer this, we need to find the potential source of ENF interference in visual recordings. We observe that the light intensity of commonly used light sources such as fluorescent and incandescent lights varies in accordance with the amount of current supplied to it. To understand the mechanism of ENF fluctuations in indoor lightings, we provide a brief background on the operation of

fluorescent and incandescent light sources.

*Fluorescent Lights:* A fluorescent lamp consists of a sealed tube containing a tungsten filament at both ends, an inert gas, and a small amount of mercury vapor. On passing electric current through this tube, electrons traveling in the form of electric current collide against the mercury atoms. If these electrons have enough kinetic energy, electrons in the mercury atoms become temporarily excited to higher energy levels. Unstable high energy levels cause electrons to return to lower energy levels, releasing photons with energy in the ultra-violet(UV) range. Since UV light is invisible to human eyes, the inner wall of the tube is coated with phosphor which absorbs UV light and emits light in the visible light range through another electron excitation process. Applying the electric current at the nominal ENF of 50 Hz/60 Hz to the tube heats the tungsten filament and generates free ions to establish a current path. As the current changes polarity at twice the rate of the nominal ENF, the current path inside the lamp turns on and off at 100/120 times per second. As a result, the lighting fluctuations illuminated by a fluorescent source is expected to be influenced by the ENF signal at the nominal frequency of 100 Hz/120 Hz.

*Incandescent Lights:* The mechanism of light production in an incandescent bulb is substantially different from a fluorescent light source [12]. Incandescent lamp technology uses electric current to heat a coiled tungsten filament to incandescence. The glass container that covers the filament is filled with a mixture of nitrogen and a small amount of inert gas such as argon. The light photons emitted by the bulb is dependent on the temperature of the coil, which depends on the heat energy dissipated in the coil and follows a power law with the amount of current

supplied. As a result, the light intensity from incandescent bulbs varies at double the frequency of the nominal ENF value and can be expected to be influenced by the ENF signal at the nominal value of 100 Hz/120 Hz, similar to the fluorescent lighting. Other commonly used electric light sources such as halogen bulbs also exhibit similar characteristics as incandescent bulbs [13].

### 2.2.1 ENF Extraction using Optical Sensors

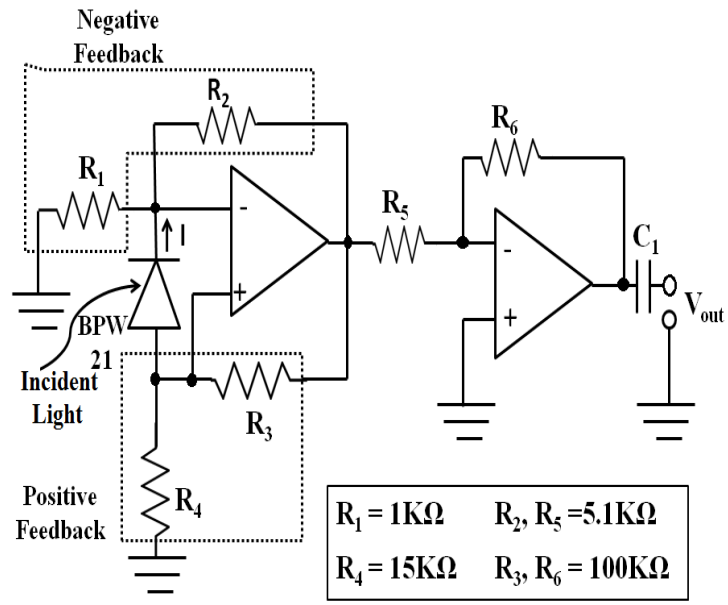
In this section, we describe the method used to extract ENF signals from fluorescent and incandescent lightings. Although the mechanism of light production in these light sources differ, ENF signal from the light intensity signal exhibits high correlation with the ENF signal captured directly from power mains.

The typical wavelengths of light emitted by fluorescent and incandescent lightings are in the visible range (400nm-700nm wavelength). To record the light intensity, we use photodiodes that produce electric current when light falls on it. We choose BPW21 photodiodes that have light sensitivity of  $9nA/lx$  and high spectral sensitivity in the visible range. Using a diode with high spectral sensitivity in the visible range ensures that infra-red signals generated from nearby heating sources do not interfere with the measurements obtained using the diode. As discussed later in Section 2.2.4, when fluorescent lamp directly above the sensor is switched off, the ENF signal is still recorded due to low ambient light present in the room where recordings are conducted. The current generated by the photodiodes is on the order of  $\mu A$ , that is too weak to be directly recorded by common current measuring

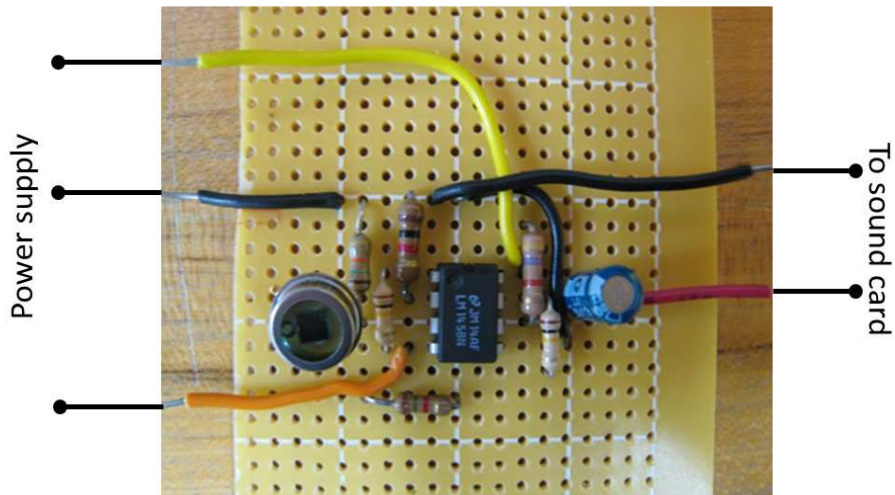
devices. We use a current-to-voltage amplifier to convert the current to measurable voltage levels. The schematic and assembled electric circuits used for this amplification purpose are shown in Fig. 2.1(a) and 2.1(b). In the first stage, a combination of a positive and a negative feedback loop is applied to obtain a robust amplification. Using the positive feedback loop increases the gain factor, producing less noisy amplification as discussed in [14]. In the second stage, a negative feedback loop is applied to further amplify the output from the first stage. For the given values of circuit parameters in Fig. 2.1(a), the overall gain of this circuit from input current  $I$  induced by incident light to output voltage  $V_{out}$  is approximately  $8 \times 10^6 \Omega$ . The output signal is given as an input to a PC sound card and sampled at 1 KHz.

To study the correlation between the ENF signal in indoor lighting captured using optical sensors and that in power mains supply, we record the ground truth ENF signal directly from the power mains of the electrical supply using a step-down transformer to convert the power supply voltage level to 5V. The transformer output is then given to a voltage divider to limit the current and voltage to safe levels to be given as input to a PC sound card. A similar mechanism was used to capture the ENF signal in audio forensics [4]. These experiments are conducted in the state of Maryland (part of the eastern power grid of the United States), and the ENF signal from lighting captured by optical sensors is expected at around 120 Hz.

To extract ENF signals, the recorded signals are given as input to an anti-aliasing low-pass filter with passband of 125 Hz. The resulting signals are down-sampled by a factor of 4 to reduce the sampling rate from 1 KHz to 250 Hz, and then passed through an equiripple bandpass filter with a narrow passband centered



(a) Circuit diagram



(b) Assembled circuit

Figure 2.1: Light sensing circuit used to verify presence of ENF signal in indoor lightings.

at 120 Hz for the optical sensor signal and 60 Hz for the power mains signal, respectively.

## 2.2.2 Instantaneous Frequency Estimation

In this section, we describe the methods used to extract and compare the ground truth ENF signal captured from the power mains with the ENF signal from indoor lightings captured using optical sensors. After applying pre-processing to recorded signals using the method described in Section 2.2.1, we estimate the dominant instantaneous frequencies in each signal to measure the fluctuations in ENF as a function of time. For this purpose, we explore different types of frequency estimators based on parametric and non-parametric spectrum estimation techniques.

### 2.2.2.1 Non-Parametric Methods

In this section, we describe the time domain zero-crossing and spectrogram based frequency estimation techniques, which do not assume an explicit model for the data.

**i. Zero-Crossing Method:** We first examine the zero-crossing method [15] to estimate the instantaneous ENF frequency. We divide the recorded signal into overlapping frames of  $T_{frame}$  seconds each with an overlap factor of 50%, and count the number of times the signal crosses zero in each frame. The dominant instantaneous frequency in each frame is recorded as half of the zero-crossing count. This method is easy to compute, but does not give a high estimation accuracy for slowly varying signals. Next, we examine spectrogram based methods that provide a high



frequency resolution in analyzing ENF fluctuations.

**ii. Spectrogram Methods:** Spectrogram is a common time frequency analysis and visualization tool. It can facilitate dominant instantaneous frequency estimation in a signal as a function of time by calculating and displaying the short time Fourier transform (STFT) of the signal [16]. To obtain the spectrogram of the ENF signal, we divide the signal into overlapping frames of  $T_{frame}$  seconds each with an overlap factor of 50%. A high resolution 8192 points FFT is taken for each frame, providing a frequency resolution of approximately 0.03 Hz for a pre-processed signal at 250 Hz sampling rate. After obtaining the spectrogram of the ENF signal, we use the following frequency domain techniques to estimate the dominant instantaneous frequencies in the ENF signal:

**(a) Maximum energy:** In this method, the frequency corresponding to the maximum energy in each time bin of the spectrogram is identified. The resulting one-dimensional signal represents the instantaneous frequencies as a function of time. This method is susceptible to frequency outliers caused by sudden changes in energy due to such factors as sudden variations in ambient lighting, etc. For example, movement near sensors can cause change in the dominant frequency.

**(b) Weighted Energy:** The weighted energy method records the weighted average frequency in each time bin of the spectrogram. The average frequency is obtained by weighing frequency bins around the nominal ENF value with the corresponding energy present. That is:

$$F(n) = \frac{\sum_{l=L_1}^{L_2} f(n, l) |S(n, l)|}{\sum_{l=L_1}^{L_2} |S(n, l)|}, \quad (2.1)$$

where  $L_1 = \lfloor \frac{(f_{ENF}-0.5)N_F}{f_s} \rfloor$  and  $L_2 = \lceil \frac{(f_{ENF}+0.5)N_F}{f_s} \rceil$ ;  $f_s$  and  $N_F$  are the sampling frequency and the number of FFT points used to compute the spectrogram;  $f(n, l)$  and  $S(n, l)$  are the frequency and the energy in the  $l^{th}$  frequency bin of the  $n^{th}$  time-frame of the recorded signal's spectrogram, respectively. Since we are interested in estimating the dominant instantaneous frequency around a known frequency value, i.e., the nominal ENF value, the value of  $l$  is chosen to include the band within  $\pm 0.5$  Hz of the ENF frequency of interest. As weighting brings robustness against outliers, this method is a more accurate estimator of the instantaneous ENF frequencies than the maximum energy method. Compared to the time-domain zero-crossing method, the spectrogram based methods for instantaneous frequency estimation provide significantly higher accuracy.

### 2.2.2.2 Parametric Methods

Subspace methods are a class of commonly used parametric spectrum estimation techniques for frequency estimation of a sinusoidal signal submerged in additive noise. These methods are primarily used in applications where the length of a given signal is small, and the separation between the frequencies to be retrieved may be shorter than the Fourier resolution limit. For our problems, ENF signals in multimedia recordings can be quite noisy and the dynamic range of the frequency variations

is often small. Subspace methods may assist us in obtaining accurate estimation of ENF signals and increasing the temporal resolution of ENF matching in such scenarios. We examine two widely used subspace based methods for ENF signal estimation: the MUSIC [17] and the ESPRIT [18].

**i. MUSIC:** The Multiple Signal Component (MUSIC) algorithm is a method used to estimate the underlying parameters of a signal from the given observations [17]. This method improves the simpler Pisarenko algorithm and estimates the frequency components of a signal consisting of a known number of sinusoids, submerged in white noise. The method relies on the orthogonality property of the sinusoidal signal subspace and the noise subspace, and provides a high resolution frequency estimate of a sinusoidal signal using a smaller number of data points than spectrogram based methods.

More specifically, the MUSIC algorithm estimates the frequency component of a signal by using an eigenspace analysis. The method estimates the  $M \times M$  auto-correlation matrix  $R_x$  of the given signal,  $x(n)$ , consisting of  $p$  complex sinusoids in white Gaussian noise. The eigen-vectors corresponding to the  $p$  highest eigen-values of the matrix  $R_x$  span the signal subspace; the remaining eigen-vectors span the noise subspace. After estimating the eigen-vectors of the matrix  $R_x$ , the frequency estimation function of the MUSIC algorithm is:

$$P_{MUSIC}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M |e^{H} \mathbf{v}_i|^2} \quad (2.2)$$

where  $\mathbf{e} = [1 e^{j\omega} \dots e^{j(M-1)\omega}]^T$  and  $\mathbf{v}_i$  represents an eigen-vector from noise sub-

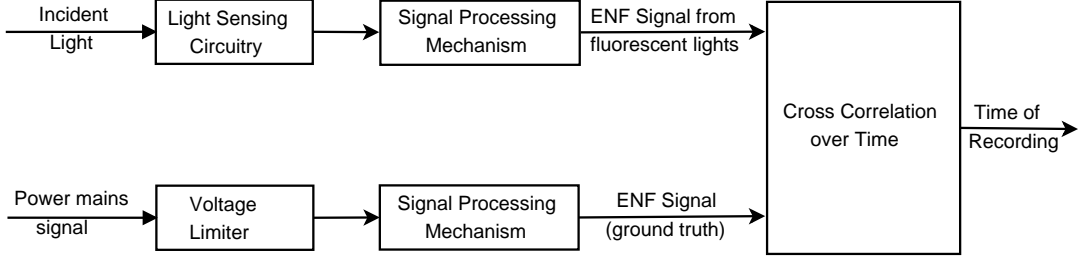
space. The  $\omega$  values corresponding to the peaks in  $P_{MUSIC}(e^{j\omega})$  are the estimated frequencies present in the sinusoid signal  $x(n)$ . An alternative version of the MUSIC algorithm, known as Root MUSIC [19], offers a high resolution to estimate the above mentioned  $\omega$  values with substantial savings in computation power.

**ii. ESPRIT:** The Estimation of Parameters using Rotational Invariant Techniques (ESPRIT) is a method similar to the MUSIC. Substantial differences between the MUSIC and the ESPRIT include that the ESPRIT uses the signal subspace while the MUSIC uses the noise subspace [18], and that the ESPRIT estimates the signal subspace from the data matrix, while the MUSIC does an explicit computation of the correlation matrix. Similar to the MUSIC, the ESPRIT is promising to provide a robust estimate of signal parameters - the frequency in our case - using a smaller number of data points as compared with spectrogram based techniques.

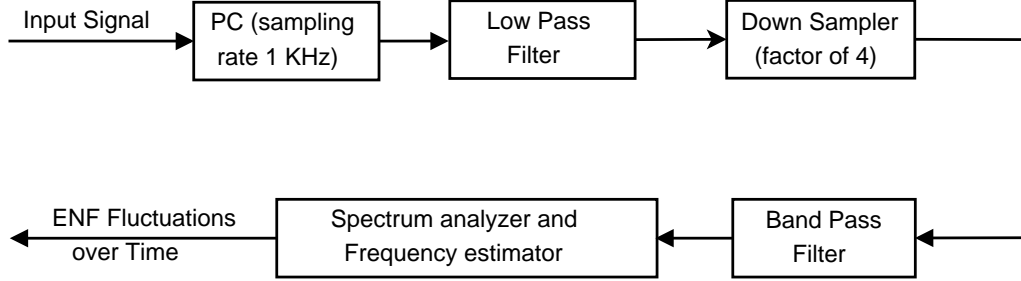
In our work, the instantaneous frequency signal, which we refer to as ENF signal for short, is estimated by dividing the given signal into overlapping segments of  $T_{frame}$  duration with an overlapping factor of 50%, and obtaining an estimate of the frequency of each segment by applying the frequency estimation algorithms described above.

### 2.2.3 Correlation Between Sensor Signal and Power Signal

After estimating the frequency fluctuations in ENF signals as a function of time, we estimate the Normalized Cross-Correlation (NCC) function between the



(a) Block diagram for ENF extraction



(b) Details of the signal processing module in (a)

Figure 2.2: Block diagram for estimating creation time of a sensor recording.

two ENF signals as:

$$\rho(k) = \frac{\sum_{n=1}^N [F_s(n) - \mu_s][F_m(n-k) - \mu_m]}{\sqrt{\sum_{n=1}^N [F_s(n) - \mu_s]^2 \sum_{n=1}^N [F_m(n-k) - \mu_m]^2}}, \quad (2.3)$$

where  $F_s(n)$  and  $F_m(n)$  are the dominant instantaneous frequencies in  $n^{\text{th}}$  time-frame of ENF signal in optical sensor recording and power mains recording, respectively;  $\mu_s$  and  $\mu_m$  are the mean values of  $F_s(n)$  and  $F_m(n)$ , respectively; and  $N$  is the total number of frames in optical sensor signal. The peak value of  $\rho(k)$  represents an estimated delay of  $k$  frames in the time-of-recording between the optical sensor signal and the power mains signal. A block diagram of the overall system used to extract the ENF fluctuations over time and compare multiple ENF signals is shown

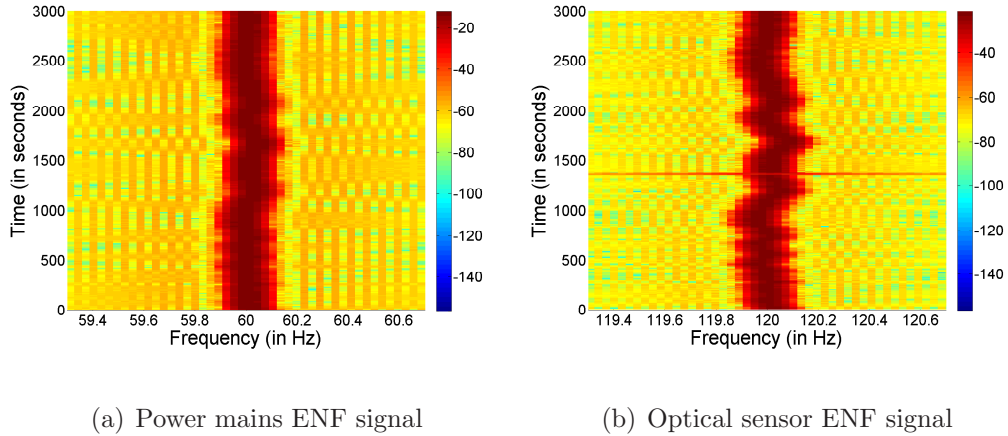


Figure 2.3: ENF fluctuations of signals captured in Experiment 1.

in Fig. 2.2.

## 2.2.4 Experimental Results and Discussions

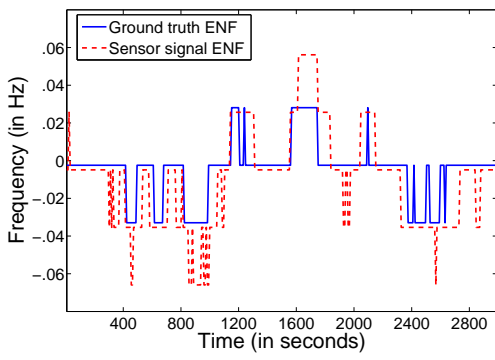
In this section, we describe different experiments conducted to investigate the presence of ENF signals under fluorescent and incandescent lightings. A light sensing circuitry shown in Fig. 2.1(b) is assembled, and its output is given as an input to a PC sound card to record it. A recording directly from power mains is also conducted in parallel to extract the ground truth ENF signal.

### 2.2.4.1 Experiment 1

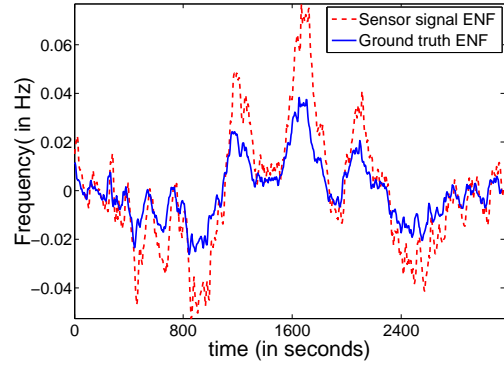
In our first experiment, a fluorescent light is switched-on near the light sensing circuitry, and data are recorded for 55 minutes. Power mains signal recording is started with a delay of 15 seconds from the optical sensor recording to determine the accuracy of our method in estimating the time-of-recording. In the controlled settings used in this and the following experiments with photodiodes, all other lights

in the room are switched-off. The spectrograms of the ENF signals obtained from the ground truth signal and the optical sensor signal around the frequencies of interests are shown in Fig 2.3(a) and 2.3(b), respectively. From these figures, we observe that ENF fluctuations captured under fluorescent lightings are similar to that of the ground truth ENF signal, and the fundamental component of the ENF signal from sensors is present at a frequency of 120 Hz. As the 120 Hz component in the sensor signal represents the fundamental component of the ENF, in order to facilitate examination of the ground truth ENF signal and the sensor ENF signal on the same scale, we subtract a value of 60 Hz and 120 Hz from the ground truth ENF signal and the sensor ENF signal, respectively. We plot the ENF fluctuations in Fig. 2.4(a)-(d) using frequency estimation methods described in Section 2.2.2. High similarity can be seen between the ENF signals obtained using different frequency estimation methods for  $T_{frame} = 16$  seconds and an overlap factor of 50%.

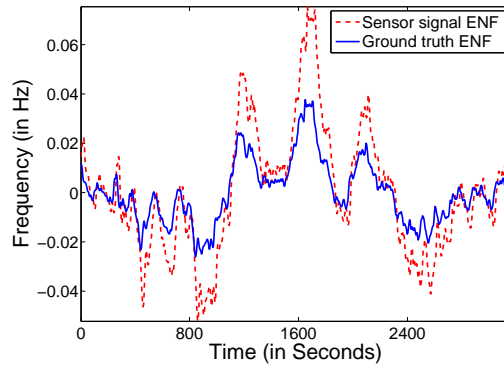
The plots of the NCC function,  $\rho(k)$ , for the zero crossing and the spectrogram based frequency estimation methods described in Section 2.2.2 are shown in Fig. 2.5(a) for  $T_{frame} = 16$  seconds. The maximum value of  $\rho(k)$  occurs at one time-frame lag for all the three methods, verifying that the delay between the sensor signal recording and power mains signal recording is in the 8–24 seconds interval. The peak value of  $\rho(k)$  found using the frequency estimates from the zero-crossing method is very small compared to those obtained by the spectrogram based methods because of the low precision of the zero-crossing method for frequency estimation. The peak value of  $\rho(k)$  obtained using the maximum energy method is smaller than that obtained using the weighted energy method. This is due to the presence



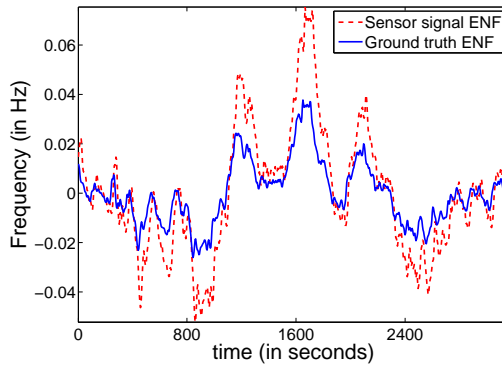
(a) Maximum Energy



(b) Weighted Energy



(c) MUSIC



(d) ESPRIT

Figure 2.4: ENF fluctuations for experiment 1 measured using different frequency domain methods for  $T_{frame}=16$  seconds.



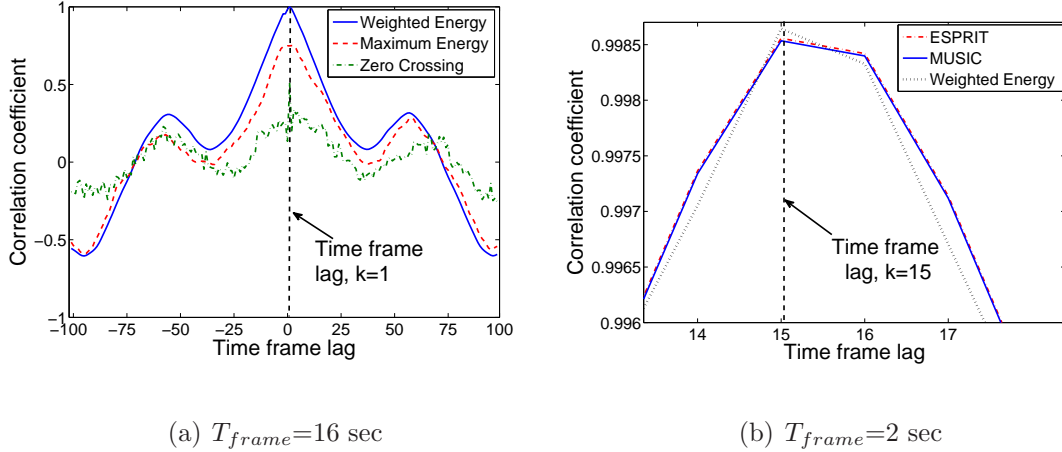


Figure 2.5: NCC coefficient,  $\rho(k)$ , for experiment 1 as a function of time-frame lag  $k$  (a) zero-crossing and spectrogram methods; (b) weighted energy and subspace methods.

of outlier frequencies that affect the frequency estimation in the maximum energy method. To achieve a better time resolution for time-of-recording estimation,  $T_{frame}$  duration can be reduced. However, decreasing the value of  $T_{frame}$  may lead to a low frequency resolution due to the time-frequency tradeoff observed in frequency estimation approaches.

In Fig. 2.5(b), we plot the NCC function for weighted energy spectrogram and subspace based frequency estimation methods described in Section 2.2.2 for  $T_{frame} = 2$  seconds. From this plot, we observe that decreasing the value of  $T_{frame}$  leads to a higher resolution in time lag estimation, as NCC peak for  $T_{frame} = 2$  is found at  $k = 15$  indicating that the start time of the recording can be narrowed down to a delay of 15–17 seconds interval. Frequency estimates from the subspace and spectrogram methods are almost same, and hence, the NCC functions shown in

Fig. 2.5(b) are almost overlapping with each other. In general, decreasing the value of  $T_{frame}$  leads to an increase in the resolution of time lag estimation. However, the peak value of correlation coefficient decreases. Decreasing the value of  $T_{frame}$  further gives lower confidence in estimation of time-of-recording as the peak value of the NCC function decreases and side-lobe value increases.

We would like to note here that signals recorded from the light sensing circuitry are of high signal-to-noise ratio (SNR). As a result, we can reduce the  $T_{frame}$  to a small value and still expect to obtain good estimates of instantaneous frequencies and achieve very fine temporal resolutions. However, for the cases of video recordings, the recorded ENF related signals are noisy due to camera sensors' low sensitivity and interference from visual content. For such cases, as will be shown in Chapter 3, the performance in timestamp estimation degrades significantly as the  $T_{frame}$  duration is decreased. Reducing the value of  $T_{frame}$  leads to noisy frequency estimates and makes the time lag estimation worse.

#### 2.2.4.2 Experiment 2

In our second experiment, a fluorescent light is switched off for the first 17-minutes and then switched on for the next 13-minutes. In this experiment, the power mains signal recording is started with a delay of 70 seconds from the sensor signal recording. The room where the experiment is performed also receives some light from a low ambient light source. The spectrogram of the optical sensor recording for this experiment is shown in Fig. 2.6. From this figure, we observe that the ENF signal is weak for the period when the light is off. However, the fluctuations

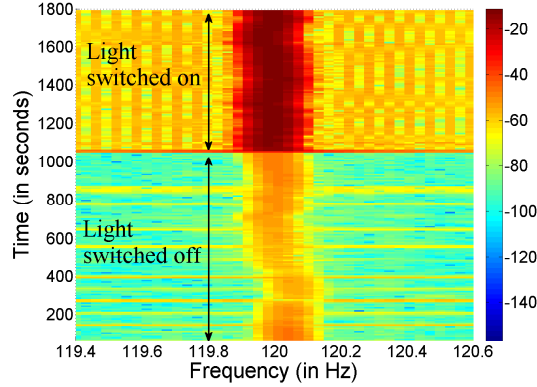


Figure 2.6: ENF for optical sensor signal in Experiment 2, including a weak signal for first 17 minutes from the ambient light, when the main light is switched off.

still show high correlation with the ground truth signal (the corresponding power mains signal not shown due to limited space), implying that optical sensors can also capture fluctuations from ambient fluorescent lights. In this case, the NCC coefficient is found to be maximum for  $k = 8$  or equivalently at a time delay of  $64 - -80$  second interval for  $T_{frame} = 16$  seconds, verifying the approximate timing at which the recording started. The temporal resolution in time lag estimation can be increased by reducing the value of  $T_{frame}$  and using the subspace based frequency estimation methods.

### 2.2.4.3 Experiment 3

In our third experiment, an incandescent lamp is switched on at a delay of 19 seconds with the power mains signal recording. The signal from the light sensing circuitry is recorded for 20 minutes. The peak value of the NCC function occurs at  $k = 18$  for  $T_{frame} = 2$  seconds and 50% overlap factor, as shown in Fig. 2.7.

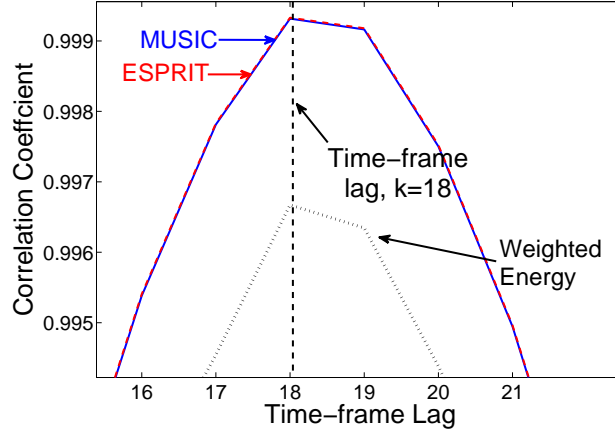


Figure 2.7: NCC coefficient,  $\rho(k)$ , for experiment 3 as a function of time-frame lag  $k$  using different frequency estimation methods for  $T_{frame} = 2$  seconds. Curves obtained using the MUSIC and the ESPRIT frequency estimation methods are closely overlapping with each other.

This indicates that the recording began at a delay of 18–20 seconds from the power mains signal. Similar to the observations in experiment 1 and experiment 2, we have observed that the number of side-lobes increases for decreasing the value of  $T_{frame}$ .

In our experiments, we also observe a weak signal around 60 Hz in the optical sensor recordings. The energy in this signal is approximately 20 times less than that of the ENF signal at 120 Hz when fluorescent or incandescent light falls on the circuit. The 60 Hz signal component is highly correlated with the ground truth ENF signal, suggesting that ENF fluctuations are also captured in the optical sensor signal around the nominal ENF value of 60 Hz. We conjecture that the source of this signal may be from electromagnetic interferences of power supply with the light sensing circuitry.

## 2.3 Chapter Summary

In this chapter, we demonstrated that the near invisible flickering of such indoor lighting environments as fluorescent and incandescent lightings is a source of ENF fluctuations in optical recordings. We used an optical sensor circuitry and conducted different experiments in fluorescent and incandescent lighted environments. We also developed a signal processing mechanism to estimate the ENF signals from optical sensor recordings and power recordings. We used spectrogram and subspace based frequency estimation methods to estimate the instantaneous frequencies in ENF signals. A correlation coefficient based matching was performed to demonstrate that the ENF signal from power recordings exhibit a high correlation with the ENF signal from optical sensor recordings at the corresponding creation time and can be used to estimate/verify the creation time of optical sensor recordings. Subspace based frequency estimation methods such as MUSIC and ESPRIT was shown to provide a better temporal resolution in creation time estimation as compared with spectrogram based frequency estimation methods.

# Chapter 3

## Video Forensics using ENF Signals

### 3.1 Chapter Introduction

The recent decade has witnessed a huge amount of multimedia data in the form of audio, image, video, etc. Such data can make rapid and broad social impacts when multimedia recordings are distributed to users through social networks. Video recordings are also increasingly used for surveillance purpose in defense and security applications. These recordings are generally stored on disks or other storage devices, and contain metadata describing important information such as the time and the place of recording. However, the digital nature of multimedia data makes it vulnerable to digital forgeries. For example, many digital edifying software tools can be used to modify the creation time/location information stored as metadata, or to cut a clip of one recording and insert it into another. These modifications can result in serious consequences when multimedia recordings are used for law enforcement and journalism cases. In the absence of any cryptographic protection

and watermarking techniques during the initial data acquisition, such modifications can be difficult to detect. Developing forensic tools which can identify data origin and detect content tampering is necessitated in views of the feasibility of digital forgeries.

As Electrical Network Frequency (ENF) fluctuations have been shown as time dependent signatures which are embedded in audio and optical sensor recordings at the creation time, it can also prove to be useful in forensic analysis of digital videos. As shown in Chapter 2, flickering in indoor lighting environment carry ENF information and can be extracted using a suitable signal processing mechanism. Most consumer end video cameras use charged coupled device (CCD) and complementary metal oxide semiconductor (CMOS) sensors, which are optical sensors capturing light coming from different places in the field-of-view of the camera. These recordings can be influenced by the light coming from electric powered indoor lighting devices such as fluorescent lamps and/or incandescent lamps.

In this chapter, we demonstrate the challenges in extracting ENF signals from video camera recordings. In particular, sensors used in video cameras capture visual data at the rate of 25/30 frames-per-second, which is considerably lower than the nominal ENF value of 50 Hz/60 Hz. A low sampling rate of such cameras introduces significant aliasing of the ENF component in video recordings, especially for videos recorded using CCD cameras. We devise a methodology that extracts the ENF signal from video recordings in indoor lighting under this aliasing. The ENF signal from optical sensor recordings and video recordings shows a significant correlation with the reference ENF signal extracted from power mains recordings, and can be

used to verify or estimate the time-of-recording. We also devise a new method to extract the ENF fluctuations from CMOS camera recordings by taking advantage of the sequential sampling of rows of a frame using the rolling shutter mechanism in these cameras [20].

As information about the time-of-recording can facilitate the integrity verification of video recordings, we examine potential applications of the proposed technique to authenticate surveillance videos. Another potential multimedia application that we study in this chapter is to determine if the audio and visual tracks in a video recording are actually captured together, or if the audio track from a different recording was manually added to the visual track. ENF signals extracted from audio and visual tracks can provide natural alignment and binding of audio-visual recordings.

## 3.2 ENF Extraction from Video

ENF signal extraction in video recordings is not as straightforward as compared with the case of optical sensors due to the low temporal sampling frequency (frame-rate) of video cameras. The nominal value of ENF is considerably higher than the sampling rates in commonly available video cameras. Most of the commercial video cameras are available at three different sampling rates in terms of frames-per-second (fps). Cameras using 24 fps are used in movie making, while most amateur hand-held cameras are available in 25 fps or 30 fps. Cameras with 25 fps sampling rate originated from PAL analog TV standard prevalent in Europe and most of Asia, while 30 fps camera originated from NTSC standard prevalent



across North America and Japan [21].

The actual frame-rate used in most video cameras may not be exactly 25 fps or 30 fps, but varies slightly. The PAL standard specify the frame-rate as 24.98 fps, while the NTSC standard is at 29.97 fps. These slight changes in the recording rates from the nominal values were introduced to make the transmission of color video compatible with the black and white television standards and the then newly introduced color televisions in the 1960s. As different digital video camera manufacturers may not strictly follow the standards, it is common to have NTSC digital cameras with a frame-rate of either 29.97 fps or 30 fps. Similarly, PAL cameras commonly have frame-rate of 24.98 fps or 25 fps. A video recording using these readily available cameras under indoor lighting may lead to significant aliasing of the captured ENF fluctuations depending on the temporal sampling frequency of the camera.

### **3.2.1 Aliasing Analysis**

ENF signals recorded by video cameras under electric powered indoor lightings experience significant aliasing due to the lower temporal sampling rate of cameras compared to the 100 Hz/120 Hz frequency components present in the light flickering. ENF signals in videos appear at different predetermined frequencies, which can be analyzed from the sampling theorem [22].

As an example, we examine the effect of a video recording in indoor lightings powered by a 50-Hz source. Since the current changes polarity at twice the power

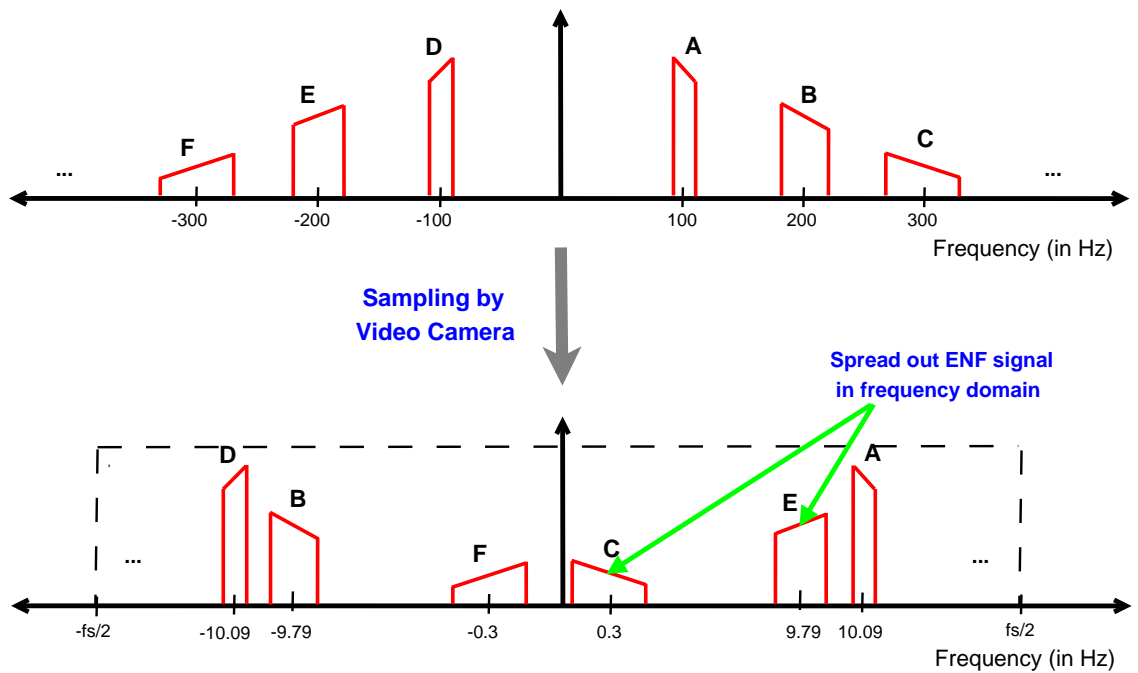


Figure 3.1: Spectral domain illustration of the aliasing effect in video capturing of ENF signals: (Top) Spectral composition of the original fluorescent light signal powered using 50 Hz supply frequency; (Bottom) Spectral composition of ENF related signals after sampling at 29.97 fps by video camera.

mains frequency, the light flickers at 100 Hz. Additionally, when the power mains signal slightly deviates from a perfect sinusoid form, higher harmonics of decaying energy are often present at integer multiples of 100 Hz. The bandwidth of these higher harmonics is also greater than the main component because the practical ENF signal of interest is a narrowband signal and not a perfectly stable sinusoid. So, the bandwidth of the  $k^{\text{th}}$ -harmonic component will be  $k$ -times the bandwidth of the main ENF component at 100 Hz. In addition, in practical visual recordings, we have observed little or no anti-aliasing lowpass filtering along temporal axis during the initial analog-to-digital acquisition in our experiments. Suppose the camera used for capturing video is an NTSC standard camera with a frame-rate of 29.97 Hz. Because of the low temporal sampling rate as compared to the required Nyquist sampling rate to avoid aliasing, the resulting spectra will have periodic tiling of frequency components at  $\pm 100 + 29.97k$ ,  $k = 0, \pm 1, \pm 2, \dots$ . Because of the periodic nature of the resulting spectra, it suffices to focus on replicas within one period. Aliasing effect introduced due to such video recordings is shown in Fig. 3.1. From this figure, we observe that multiple copies of the ENF related components will appear in the temporal spectrum of the video signal around different but pre-determined frequencies. These multiple copies arise due to the presence of higher harmonics in the power mains signal, and may be combined strategically to obtain a better estimate of the video-ENF signal [23].

We further recall that the magnitude spectrum of a real valued signal is symmetric about the y-axis. Because of this, the original spectrum of the ENF signal in indoor lighting also has symmetric components at  $-100$  Hz and its harmonics,

Table 3.1: Aliased frequency for different combinations of power mains frequencies and video camera frame-rates.

Power Mains (Hz)	Video Frame Rate (fps)	Aliased Base Frequency (Hz)	Aliased 2nd Harmonic (Hz)
60	29.97	0.12	0.24
60	30	0	0
50	29.97	10.09	9.79
50	30	10	10

as shown in Fig. 3.1. After sampling by a video camera, the frequency component E, which is at  $-200$  Hz in the spectrum of the original signal, appears at 9.79 Hz in the recorded signal; and the component A that was present at 100 Hz in the spectrum of the original signal appears now at 10.09 Hz. As a result, we obtain replicas of the ENF signal at 9.79 Hz and 10.09 Hz, and they are mirrored versions of each other with different bandwidths. Similar analysis can be performed on other combinations of different camera frame-rates and power mains frequencies to find the frequencies at which the main component and the second harmonic component of the signal appear. The values of these frequencies are summarized in Table 3.1.

### 3.2.2 Experimental Setup for Video Recordings

In this section, we describe the settings under which experiments are conducted to detect the presence of ENF signals in videos. We have collected video data from three geographical regions – China, India, and USA – to demonstrate the capability of the ENF signal to act as a timestamp and authentication signal. For each of these areas, we record videos under the following two settings:

### 3.2.2.1 White Wall Video

In this setting, we record a constant white wall scene illuminated under fluorescent lights for 10 minutes and examine the video for ENF traces. This experiment can be considered a first step towards demonstrating the presence of ENF signals in videos recorded under fluorescent lightings. As incandescent lightings were also shown to be influenced by the ENF signal in Chapter 2, similar results to those reported in this section can be expected for video recorded in incandescent lightings.

To extract the ENF signal from video recordings, we compute the average intensity of each frame, and pass this intensity signal over time through a temporal bandpass filter with passband corresponding to the frequencies of interest, as listed in Table 3.1, where aliased components of the ENF signal are expected to appear in videos. As the content of the video is mostly the same in each frame, the presence of a significant amount of energy in the frequency of interest can be attributed to the ENF signal. For all the experiments conducted on video data, the frequency estimation algorithms discussed in Chapter 2 are applied to extract the ENF signal around the aliased frequency bands.

### 3.2.2.2 Surveillance Video

In this setting, a camera is placed inside a room for surveillance purposes; the camera locations vary in several recording sessions. Fluorescent lights illuminate the room. Occasional movement of people can be seen in the foreground of the video. Such recordings can be considered representative of real-life surveillance scenarios

in places where few events are expected to occur. In these experiments, because the content in each frame of the video is not constant, a direct averaging of the frame pixel values may not be a suitable preprocessing before performing the frequency analysis. As ENF signal from fluorescent lightings is recorded by each sensor pixel in the form of light intensity interferences, we can use a part of the video frame to extract the ENF signal. For example, we locate the relatively stationary areas in video frames where content does not change much, mostly at the top and bottom areas of video frames in our experiments. We use these areas to extract the ENF signal based on the spectrogram of average pixel intensity. Alternatively, the process of finding the static regions in frames of the video can be automated using common motion estimation techniques. In the next section, we will report the experimental results under these settings for forensics applications.

### **3.3 Applications of ENF Signal Analysis to Video Forensics**

In this section, we describe some potential applications of the ENF signal analysis in video forensics. We examine the use of the proposed technique to estimate or verify the time-of-recording, detect clip insertion/deletion operation in videos, and forensically bind audio-visual data to determine if the audio and visual tracks of a video were recorded at the same time or superimposed from different recordings.

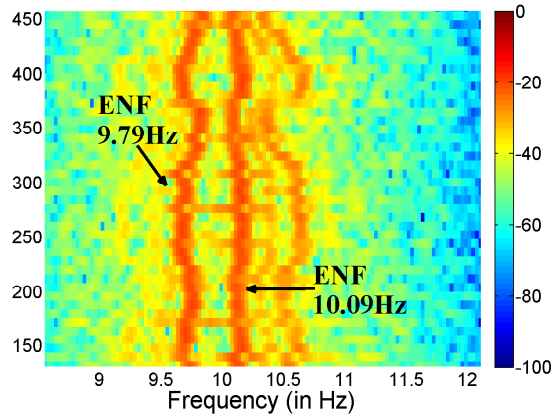


Figure 3.2: ENF fluctuations measured for the whitewall video experiment in China.

This figure is best viewed in color.

### 3.3.1 Estimating or Verifying Time-of-Recording

To study the applications of ENF signal analysis for estimating and verifying the time-of-recording, we record several videos with a known time lag from the power mains signal. ENF signal is extracted from these recordings, and the NCC function is calculated between the ENF signal from each video recording and that from the power mains. The time corresponding to the peak value of the NCC function is the estimate of the time lag between video and power mains recording. In a large database of power mains frequency data with timestamp information, the peak value will suggest the actual time-of-recording of the video data. We now present results of timestamp estimation experiments conducted in different geographical areas.

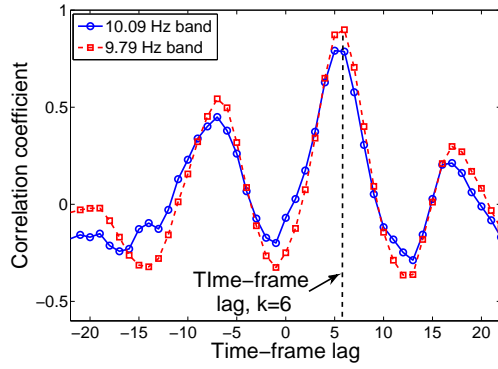
### 3.3.1.1 Recordings from China

In China, 50 Hz supply frequency is used in power distribution networks. As a result, the ENF signal in the light flickering is around 100 Hz. We use a Panasonic camera with a frame-rate of 29.97 fps to record videos in China. For this experiment, video recording is started at a delay of 50 seconds with respect to the power mains signal, and the video is recorded for a duration of 10 minutes. As discussed in Section 3.2.1 and shown in Fig. 3.1, we expect the main component of the ENF signal in video recordings to be present in the band around the aliased frequency of 10.09 Hz. Higher harmonics of the ENF signal in video recording are present in the band around 9.79 Hz (from the second harmonic) and 0.30 Hz (from the third harmonic).

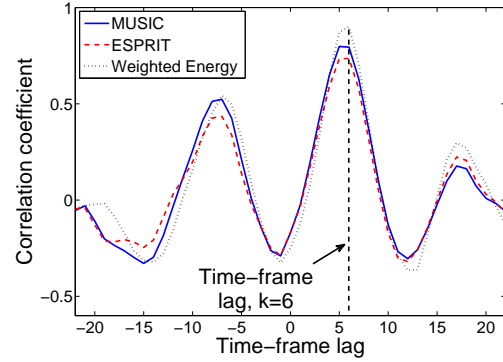
The magnified spectrogram of the ENF signal for the white wall video recording around 10 Hz frequency band is shown in Fig. 3.2, in which we can see multiple copies of the ENF signal around different frequencies in accordance with our analysis in Section 3.2.1. We also observe from this figure that the bandwidth of the video-ENF signal present around 9.79 Hz is approximately double of the bandwidth of the video-ENF signal around the frequency band at 10.09 Hz. This can also be explained from Fig. 3.1, as the copy of the video-ENF signal at 9.79 Hz originates from the second harmonic of the base ENF signal, present at 200 Hz from the light flickering.

As explained in Section 3.2.1, the 9.79 Hz component of the recorded signal is a mirrored version of the 10.09 Hz component. So, we flip the ENF signal extracted

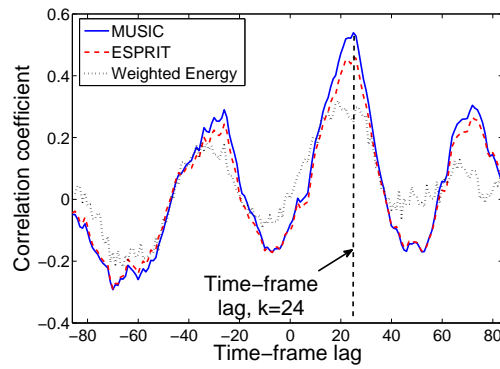




(a)  $T_{frame} = 16$  seconds



(b)  $T_{frame} = 16$  seconds



(c)  $T_{frame} = 4$  seconds

Figure 3.3: NCC coefficient,  $\rho(k)$ , as a function of time-frame lag  $k$  using different frequency estimation methods on video data from China: (a) Weighted energy; (b), (c) subspace methods on signals at the 9.79Hz band at longer and shorter time-frame size, respectively.

from 9.79 Hz component in the NCC function calculation to make the overall correlation positive. The plot of the NCC function between the video-ENF signal and power-ENF signal evaluated using the video-ENF signal present around 10.09 Hz and 9.79 Hz frequency band is shown in Fig. 3.3(a) for  $T_{frame} = 16$  seconds by the spectrogram based weighted energy method. The plot obtained using the video-ENF signal from 10.09 Hz band is similar to that obtained using the ENF signal from 9.79 Hz frequency band. The maximum value of the NCC function is obtained at a lag of  $k = 6$  time-frames. This indicates that the video recording began at a delay of 48–64 seconds interval from the power mains signal; the actual delay of 50 seconds lies in the interval. As the value of the peak correlation coefficient for the ENF signal estimated from the 9.79 Hz band is slightly higher than the 10.09 Hz band, we use the 9.79 Hz frequency band to estimate the ENF signal in our subsequent results. We plot the NCC function for the subspace and the weighted energy spectrogram methods in Fig. 3.3(b) and 3.3(c) for  $T_{frame} = 16$  seconds and  $T_{frame} = 4$  second, respectively. From these figures, we observe that for  $T_{frame} = 16$  seconds, the NCC function reaches similar peaks by both spectrogram and subspace methods at time lag  $k = 6$ . For  $T_{frame} = 4$  seconds, the peak value of the NCC function for MUSIC method is observed at  $k = 24$ , indicating that the recording was done at a refined delay estimate of 48–52 second. The spectrogram based method gives worst performance among the three methods for this case.

The plots of the peak NCC value corresponding to the time-of-recording for different values of  $T_{frame}$  are shown in Fig. 3.4. From this plot, we observe that as  $T_{frame}$  increases, the peak correlation coefficient value generally increases. It

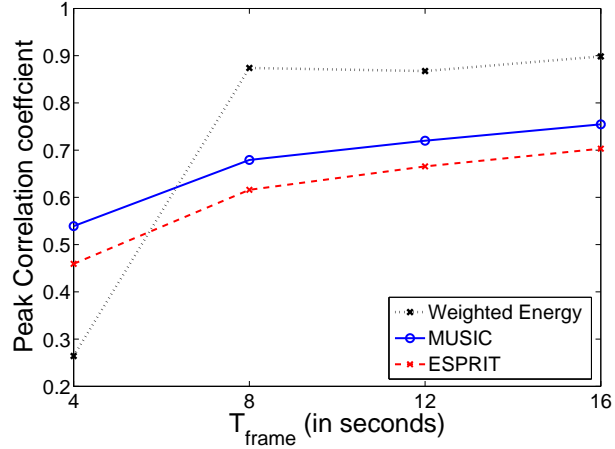


Figure 3.4: Peak NCC for different values of  $T_{frame}$  for frequency estimation at 9.79 Hz band.

is possible to extract other harmonic copies of ENF traces in the video signal by passing the video-ENF signal through bandpass filters with passband frequencies around which each harmonic copy is located. These copies of the signal can then be combined to perform correlation analysis to find the time of recording [23].

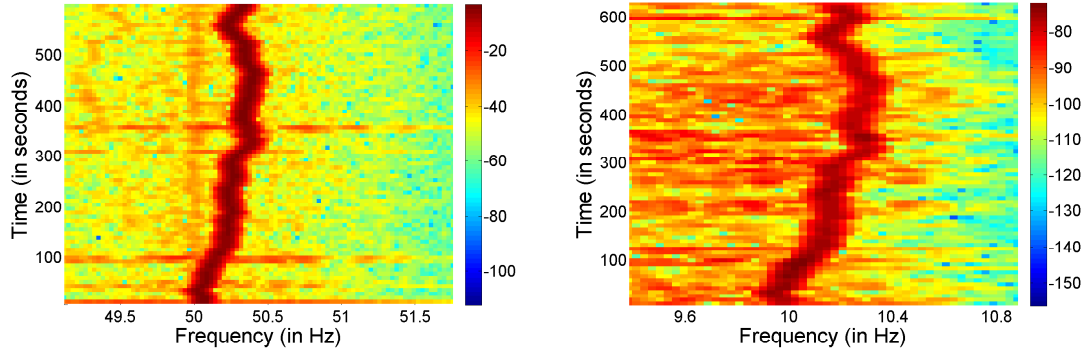
### 3.3.1.2 Recordings from India

Similar to China, the nominal value of the ENF in India is 50 Hz, hence, the ENF signal from light intensity is expected at a fundamental frequency of 100 Hz. In this experiment, we use a 30 fps NTSC Sony HDR SR12 camera to record a video for 10 minutes. The video camera recording and the power mains recording start simultaneously. From the aliasing discussion in Section 3.2.1, with this sampling frequency, the ENF signal at 100 Hz will appear at an aliased frequency of 10 Hz in video recordings. Spectrogram of the ENF signal for a white wall video recording

is shown in Fig. 3.5(b). From this figure, we can clearly observe a high correlation between the pattern of the ENF signal at 10 Hz and the ground truth ENF signal at 50 Hz obtained directly from power mains supply, as shown in Fig. 3.5(a). The maximum value of the NCC coefficient,  $\rho(k)$ , between the video-ENF signal and the power mains signal is found to be at  $k = 0$  for  $T_{frame} = 16$  seconds, as shown in Fig. 3.5(c). This indicates that the recording began at a delay of 0–16 second interval for the power mains signal and the video signal. We can obtain better time resolution in our estimation by reducing the value of  $T_{frame}$ , although the peak correlation coefficient value would generally be smaller.

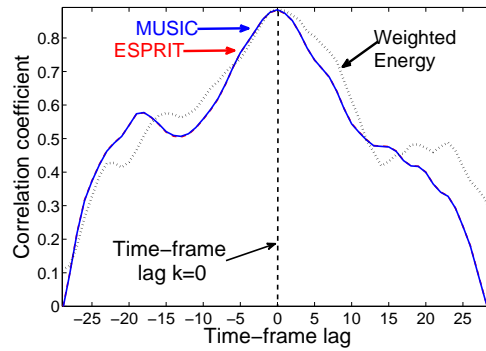
In Fig. 3.6, we plot the average peak correlation coefficient obtained over the 10 video recordings. Each of the recordings is about 10 minutes' long for the cases when ground truth ENF from each recording is matched with the corresponding ENF signal (the matching case) and when video-ENF signal from each recording is matched with the ground truth ENF signal from other recordings (the non-matching case). From this figure, it can be observed that as the value of  $T_{frame}$  increases, the average peak correlation coefficient for all the three frequency estimation method improves for the matching case. For the non-matching case, the value of the average peak correlation coefficient is close to zero, indicating that the video and the power signals were recorded at different times. We also observe that the subspace based MUSIC performs substantially better than the weighted spectrogram energy approach.

For the experiment on surveillance video recording, the first five minutes of video are recorded with the camera constantly panning such that all parts of the



(a) Power-ENF signal

(b) Video-ENF signal



(c) NCC coefficient,  $\rho(k)$

Figure 3.5: (a), (b) ENF fluctuations measured for the whitewall video experiment in India; (c) NCC coefficient,  $\rho(k)$ , as a function of time-frame lag  $k$  for  $T_{frame} = 16$  seconds.  $\rho(k)$  curves obtained using the MUSIC and the ESPRIT frequency estimation methods are closely overlapping with each other.

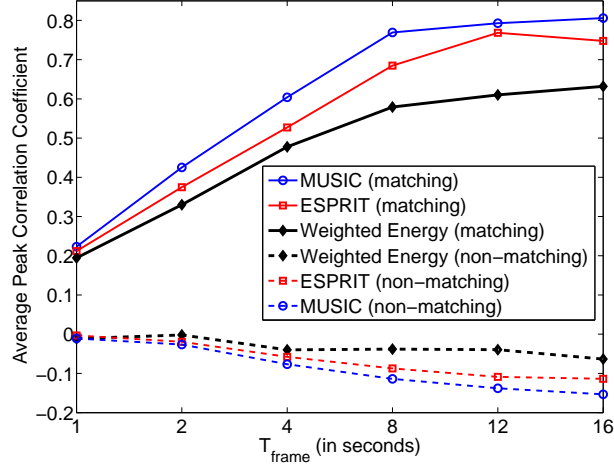
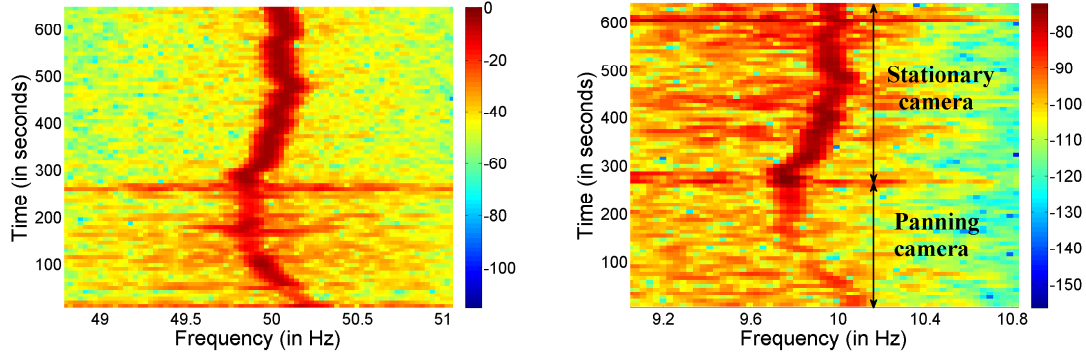


Figure 3.6: Average Peak NCC value for matching and non-matching case over 10 videos.

video frames undergo some changes, while the remaining part is recorded with the camera stationary. From the spectrogram of this recording’s ENF signal shown in Fig. 3.7(b), we see that the ENF signal is not strong during the first five minutes when the camera pans. Nevertheless, the pattern in the ENF signal around 10 Hz for the first five minutes follows a similar pattern to the power-ENF signal shown in Fig. 3.7(a). The NCC function,  $\rho(k)$ , attains the highest value at  $k = 0$ , indicating that the video signal recording began at a delay of 0–16 seconds with the power mains signal. Higher temporal resolution can be obtained by reducing the value of  $T_{frame}$  and using the subspace methods for frequency estimation.

### 3.3.1.3 Recordings from USA

The nominal value of the power supply frequency in the United States is 60 Hz. The ENF signal from the light flickering can be expected at 120 Hz. We use the

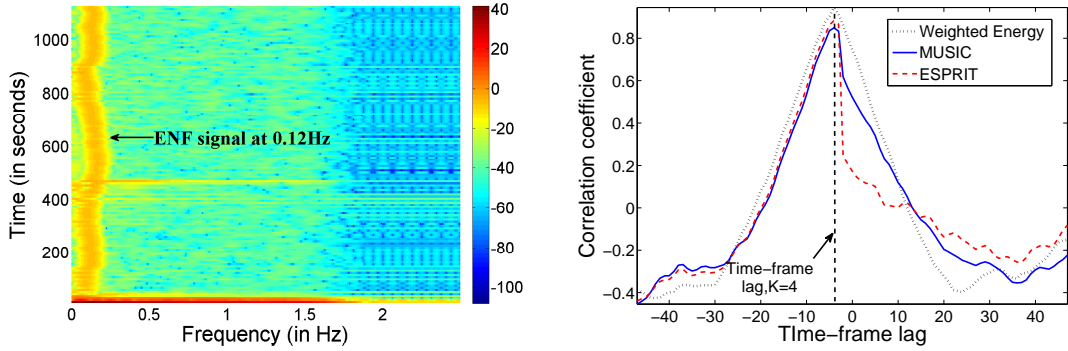


(a) Power-ENF signal

(b) Video-ENF signal

Figure 3.7: ENF fluctuations measured for surveillance video experiment in India.

same 29.97 fps Panasonic camera from experiments done in China to conduct video recording in the United States. We recall from Table 3.1 that the ENF signal in the video recordings is expected at an aliased frequency of 0.12 Hz. Fig. 3.8(a) shows the spectrogram of the ENF signal for the white wall video. This spectrogram is plotted for the time-frame duration  $T_{frame} = 16$  seconds with a 50% overlap and 1024 point NFFT. Significant energy band at 0.12 Hz confirms the presence of the ENF signal in these video recordings. The NCC function between the video-ENF signal and the power-ENF signal is plotted in Fig. 3.8(b) for  $T_{frame} = 16$  seconds. The maximum value of the NCC coefficient is found to be 0.95 corresponding to a lag of  $k = 4$ , indicating the video recording was started with a delay of 32–48 seconds from the power signal. The actual delay in recording time is approximately 38 seconds. This again shows the timestamping capability of the ENF signal in video recordings. High temporal resolution can be obtained by decreasing the value of  $T_{frame}$ . For example, the time lag estimate of  $k = 9$  is found for  $T_{frame} = 8$  seconds, indicating that recording was started at delay of 36–44 seconds interval



(a) Video-ENF signal

(b) NCC coefficient,  $\rho(k)$  for  $T_{frame} = 16$  seconds

Figure 3.8: (a) ENF fluctuations measured for the whitewall video experiments in US; (b) NCC coefficient,  $\rho(k)$ , as a function of time-frame lag  $k$ .

with the power mains.

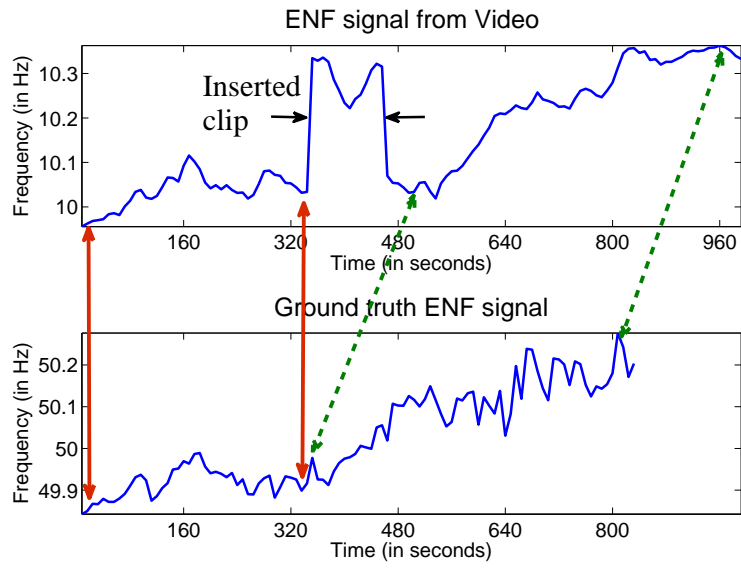
### 3.3.2 Tampering Detection

We design two experiments to demonstrate an application of the ENF signal analysis for tampering detection of surveillance videos. For this purpose, we record a 980-second video in India in an indoor environment illuminated with fluorescent lights. We perform a simple video clip insertion by cutting the last 160 seconds of video and inserting it in between the remaining clip of the video close to the 340<sup>th</sup> second. ENF signal extracted from this tampered video and the ground truth ENF signal extracted from the power mains signal database corresponding to the time-of-recording are shown in Fig. 3.9(a). From this figure, we can locate the regions in the ENF signal from the video that have similar patterns as the ground truth ENF

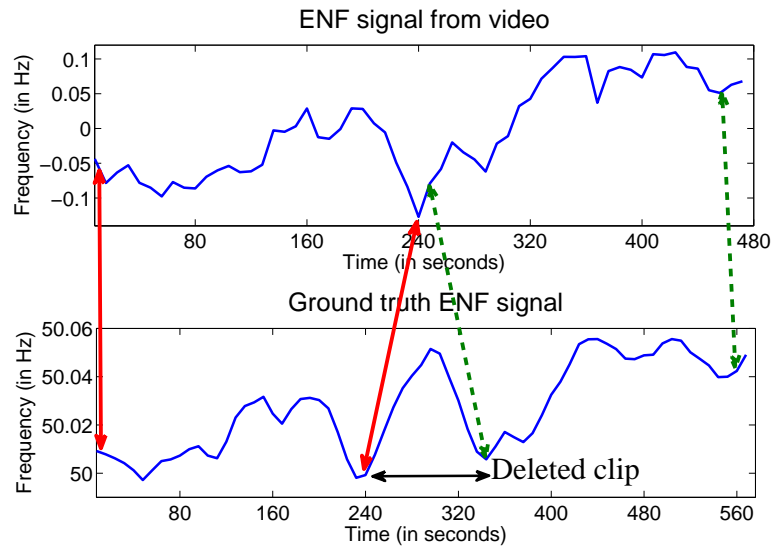


signal for the given time-of-recording. We observe that the ENF signal from video recording shows the same pattern as the ground truth ENF signal for 340 seconds in the beginning of the video as shown by the red arrows in the figure. Similarly, video-ENF signal from time index 500 second to 980 second have the same pattern as the ground truth ENF signal between time index  $340^{th}$  and  $820^{th}$  seconds. The video-ENF signal between time index  $340^{th}$  seconds and  $500^{th}$  seconds does not match with the ground truth ENF signal, suggesting that the video recording is likely to be modified in the corresponding region and a video clip insertion operation was performed. Beyond visual examinations, more rigorous verification can be performed by quantitatively comparing the segments of ENF sequence from video with the reference ENF database.

Similar to the clip insertion example, we perform a clip deletion operation on a video recorded in China. The video is 570 seconds long and part of the video clip is deleted between  $240^{th}$  second and  $340^{th}$  second. The ENF signal extracted from the tampered video and the corresponding ground truth ENF signal are shown in Fig. 3.9(b). From this example, we see that the ENF traces from the tampered video match with the ground truth ENF signal for the first 240 seconds. The remaining part of the ENF traces from the tampered video match with the ground truth ENF signal after  $340^{th}$  seconds, confirming that the video clip between time index of  $240^{th}$  second to  $340^{th}$  second is removed from the video. From these two examples, we can see that the ENF signal analysis is potentially a powerful technique to detect video tampering for videos recorded in indoor lighting.



(a) Clip insertion

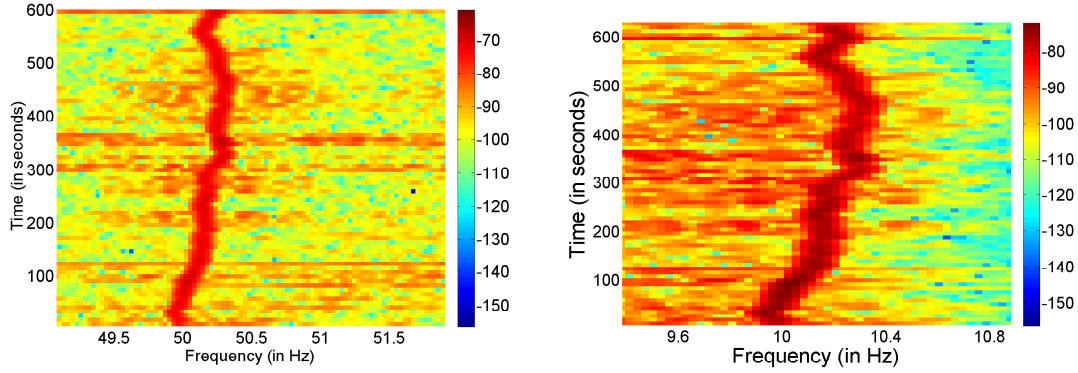


(b) Clip deletion

Figure 3.9: ENF matching result demonstrating video tampering detection based on ENF traces.

### 3.3.3 Audio-Visual Authentication and Synchronization

We now explore an important application of the ENF signal analysis for *forensic binding* of audio-visual recordings. Such binding determines whether the audio and visual tracks in a given video recording are captured together, or if the sound track from a different recording is manually placed with the visual track. ENF signals extracted from both visual and audio tracks can provide natural alignment and binding of audio-visual recordings. As noted above, ENF signals are captured in visual track of a video due to light flickering and in the audio track through electromagnetic interference and related acoustic vibrations. When audio and visual tracks are recorded simultaneously, we expect that the ENF traces in these two streams should exhibit similar patterns, and the synchronization between the audio and the visual tracks can be verified by quantitatively comparing these two signals. Discrepancies between the ENF traces in the audio and visual track would provide a strong indicator of some kind of tampering on the video recording. For example, if the audio and visual track of a given video do not contain similar ENF fluctuations, it can be claimed that the two tracks were likely not recorded together. Therefore, synchronization can be authenticated even if the ground truth ENF signal from power mains is not available at the time-of-recording. Fig. 3.10 shows an example of the ENF traces obtained from the audio and visual tracks of a video recording. From this figure, we see that the ENF signal captured in audio, as shown in Fig. 3.10(a), exhibits similar patterns and have a high correlation with the ENF signal captured in the visual track, as shown in Fig. 3.10(b).



(a) ENF signal from the audio track

(b) ENF signal from the video track

Figure 3.10: ENF matching between audio and video tracks of a video recording.

## 3.4 Discussions and Extensions

### 3.4.1 Effect of Compression

ENF signals in digital video recordings are very weak signals and can be distorted from such lossy compression as the MPEG standard family. To study the robustness of ENF signals against different compression levels, we take a video recording from China and compress it at different bit rates. The plot of the peak NCC value at the time lag corresponding to the time-of-recording for different bit rates of video compressed using MPEG-2 standard for  $T_{frame} = 16$  seconds is shown in Fig. 3.11. From this figure, we observe that the value of peak correlation coefficient decreases as the video bitrate decreases; and the MUSIC and the ESPRIT methods perform better at low bit rates as compared with the spectrogram based method. This is because the SNR of the ENF signal in the video decreases as bitrate decreases, and at lower SNR, subspace based methods have shown to provide a

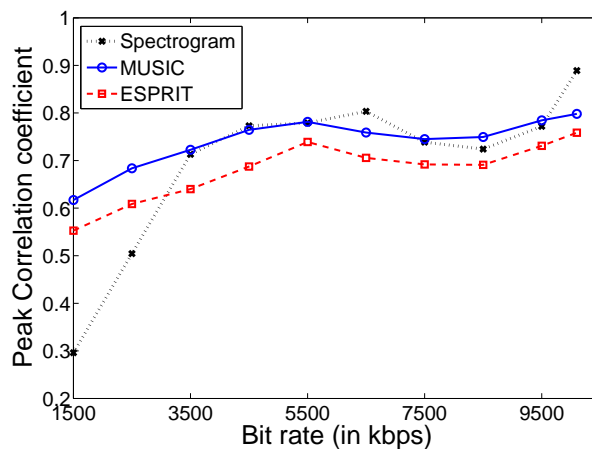


Figure 3.11: Peak NCC value for different compression rates for videos recorded in China.

better estimate of the ENF signal as compared with the spectrogram based methods.

### 3.4.2 ENF Extraction from CMOS Imaging Cameras

In this section, we explore ENF signal extraction from cameras with CMOS imaging sensors, which are increasingly used in a wide variety of imaging systems of still images and videos, such as mobile phone cameras, web cameras, and standalone digital cameras. Most consumer-end cameras equipped with CMOS sensors contain column-parallel readout circuits. These circuits read all the pixels in a row at the same time. The readout proceeds from a row to another in a sequential manner from top to bottom, with no overlap in the readout time of different rows. Such a sampling mechanism by the imaging sensors is referred to as a *rolling shutter* mechanism [20]. An example of the CMOS imaging sensor sampling mechanism is shown in Fig. 3.12.

Traditionally, rolling shutter has been considered detrimental to the image quality, as each row of the image is exposed to the light at a different time, which creates artifacts for fast moving objects in the scene. However, as each frame of a video captured using rolling shutter cameras has undergone space-time sampling, recent literature has exploited this mechanism for high-speed photography and optical flow based applications [20], kinematics, and object pose estimation [24] [25]. The spatial-temporal sampling nature of the rolling shutter provides us with a potentially high sampling rate of the ENF signal as compared to the traditional CCD-based cameras. As we recall from our aliasing analysis in Section 3.2.1, the ENF signal captured in recordings using CCD camera with 29.97 fps and 30 fps in the 60 Hz power frequency region is expected to appear at 0.12 Hz and DC frequency, respectively. ENF signal at these aliased frequencies can be obscured by the content frequency of the video recordings, making ENF extraction difficult from such recordings. By exploiting high temporal sampling on a line by line basis from CMOS camera recordings, we can mitigate the aliasing problem on the frame level.

To extract ENF signals from videos recorded using CMOS cameras, we use the spatial average of each row as a sample. We pass the resulting signal through a band-pass filter in the ENF frequency band of interest. We downsample the bandpassed signal to 1000 Hz sampling frequency, and resultant signal is used for frequency estimation using the methods described in Section ??.

We use a Canon SX230-HS camera that employs a rolling shutter mechanism in the CMOS image sensor to record 20 videos in the US, each being 10 minutes long. The frame-rate of this camera is 30 fps. We also record the reference power signal

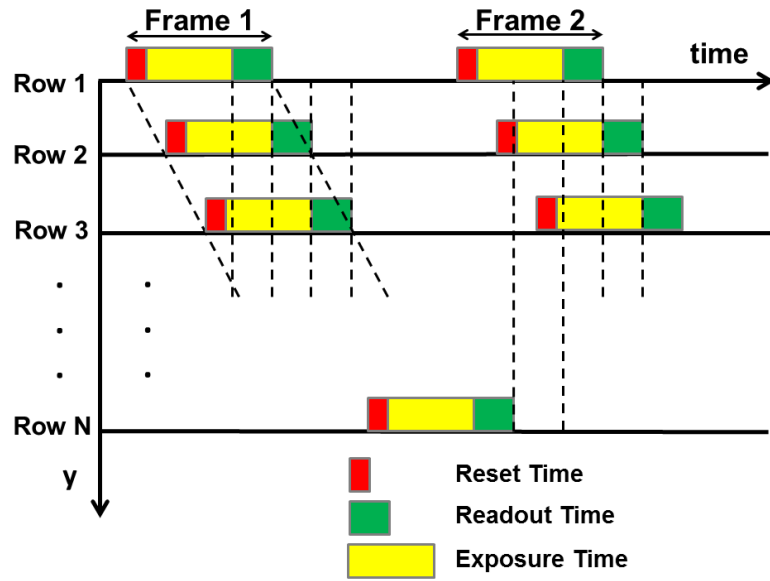
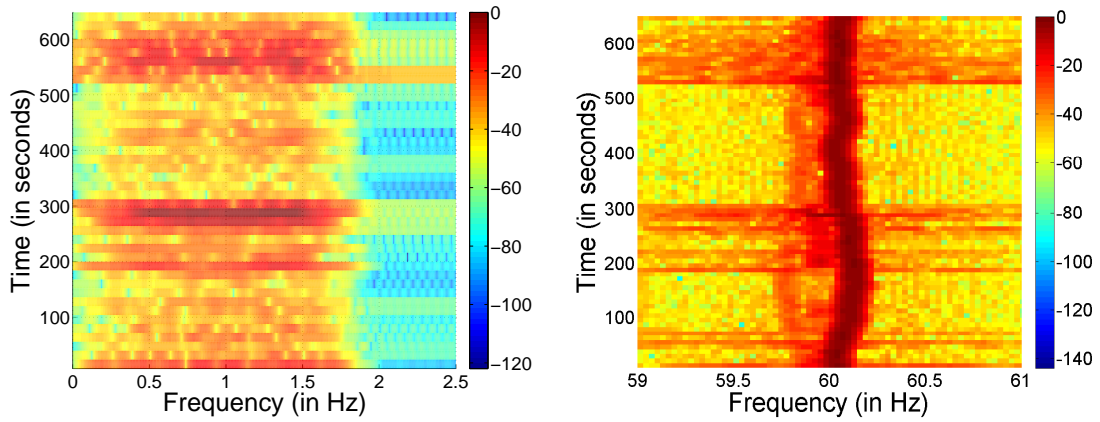


Figure 3.12: Rolling shutter sampling mechanism in CMOS cameras. Figure is adapted from [20].



(a) Whole frame signal extraction

(b) Row-by-row signal extraction

Figure 3.13: Comparison of the spectrograms of a video signal recorded using a CMOS camera and processed in two ways: a) treating each video-frame as a single sampling instant (similar to CCD mechanism), b) treating each row of a video-frame as a single sampling instant (rolling shutter mechanism).

from power mains. In Fig. 3.13(a), we plot the spectrogram of one of the recorded videos, whereby the ENF signal is processed on a frame-by-frame basis similar to the CCD camera processing discussed in Section 3.2.2. From this figure, we observe that the traces of the ENF signal are not clearly present. Fig. 3.13(b) shows the spectrogram for the same recording, when signal is processed on a row-by-row basis by explicitly considering the rolling shutter sampling. From this figure, we can clearly observe the presence of ENF signal at 60 Hz. These examples illustrate the advantages of making use of the rolling shutter effect of CMOS camera recordings in ENF signal extraction as compared with the CCD camera recordings.

In Fig. 3.14, we plot the value of average peak correlation coefficient between video-ENF signals extracted using the CMOS extraction method and the corresponding ground truth ENF signal for different values of  $T_{frame}$  and for different frequency estimation methods. From this figure, we observe a similar nature of results to what was observed in our previous results over different  $T_{frame}$  and frequency estimation methods: the correlation in the matching analysis improves for higher values of  $T_{frame}$  for a given frequency estimation methods, with subspace-based methods giving higher correlation values than the spectrogram-based method. In all of these recordings, if the whole frame is considered as one sampling instant, we do not obtain any meaningful signal by applying our ENF estimation methods. From this result, we conclude that CMOS image sensor cameras employing rolling shutter mechanism provides a better embedding of the ENF signal, and it is easier to extract these signals from recording using CMOS cameras than with CCD cameras.



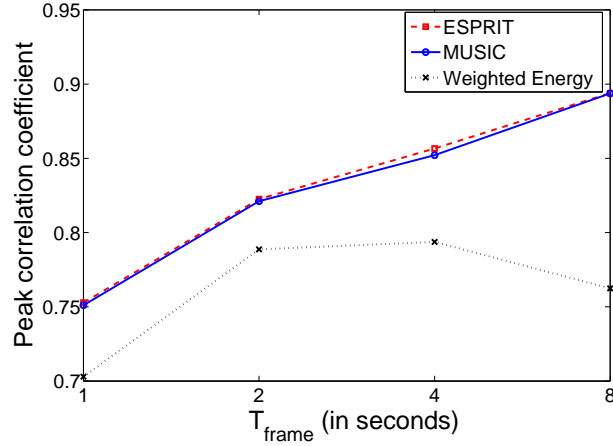


Figure 3.14: Average peak NCC for different values of  $T_{frame}$  for the videos recorded using CMOS image sensor cameras utilizing a rolling shutter based sampling mechanism.

### 3.5 Chapter Summary

In this chapter, we have demonstrated the presence of ENF signals in indoor video surveillance recordings. Video recorded under indoor lighting environment have been shown to have high correlation with ENF signals captured directly from the power mains supplies. The estimated time-of-recordings for video signals have been shown to be accurate up to a time resolution of four seconds for video-ENF signals which are noisy in nature. Subspace based methods have been shown to provide better results as compared with spectrogram based methods for low SNR ENF signals and short duration video clips for most test cases. Results from our investigations suggest that ENF signals can be used as a natural timestamp for video surveillance recordings in an indoor environment. ENF signals from video clips have

been shown to be quite robust to strong compression up to a bit rate of 250Kbps for standard size videos. ENF signals can also be used to facilitate the authentication video data, as data tampering by adversaries are likely to cause abrupt changes in the ENF signal at the corresponding part of the recordings. ENF traces also provide a natural binding of audio and visual tracks to verify their temporal synchronization and integrity.

# Chapter 4

## Statistical Modeling and Analysis of ENF Signals

### 4.1 Chapter Introduction

In previous chapters, ENF based forensics analysis was shown to be useful in authenticating multimedia recordings for time-of-recording estimation/verification, tampering detection, etc. The methodology for time-of-recording estimation using ENF signal analysis requires extraction of ENF signals from a given audio or video by means of a temporal bandpass filtering, followed by instantaneous frequency estimation as a function of time. These estimated frequencies are compared with ENF databases containing historic measurements. The time corresponding to the maximum similarity between the ENF signals from the multimedia recording and the database is the estimated time-of-recording of the given multimedia.

The techniques for the ENF signal analysis in the literature typically require a

substantial duration of multimedia recording to measure the similarity between the extracted ENF signal with the ENF database, as ENF patterns may exhibit self-similarity over time. The self-similarity may provide a wrong estimate of time-of-recording, if the ENF signal in question is not sufficiently long. This pseudo-periodic nature of ENF patterns is due to the cyclic nature of power demand and supply, and the control mechanism used to regulate power grids [6]. An increase in the load on a grid causes the supply frequency to drop temporarily; the control mechanism senses the frequency drop and additional power is drawn from adjoining areas to compensate for the increased demand. As a result, the load in adjoining areas also increases and the supply frequency drops. A similar mechanism is used to compensate for an excess supply that results in a surge in the supply frequency. In forensic applications of the ENF analysis, the effect of such pseudo-periodic patterns reduces if multimedia signal is of sufficiently long duration, typically 10-15 minutes [26]. Developing tools to authenticate relatively short recordings may enhance the benefit and affect of ENF based forensic techniques.

The first step in understanding the properties of ENF patterns requires statistical study of the signal. In this chapter, we enact a statistical study of the properties of ENF signals. Based on these properties, we model ENF signal as an autoregressive (AR) process using the Box-Jenkins methodology, which is a common mechanism to model time-series signals in forecasting and control applications [27]. We then use this proposed model to examine the performance of ENF based forensics and temporal alignment technology under a simple binary hypothesis detection framework and to understand the effect of clip duration and signal-to-noise ratio

(SNR) on the ENF matching. The application scenario considered here is timestamp verification, where the authenticity of the attached timestamp in the metadata of a given recording needs verification. Based on the proposed AR model of the signal, we decompose the ENF signal into two components: a *predictive* process and an *innovation* process. As innovation processes are uncorrelated over time, we propose the use of a decorrelated *innovation* process for ENF matching. As will be shown later, the decorrelation based matching help reduce the false alarm rate in timestamp verification [28]. We experimentally validate the proposed approach of ENF matching for timestamp verification on a 25-hours long power-audio dataset.

The proposed approach to matching ENF sequences based on innovation process may also prove beneficial for forensic analysis of a possible forgery on ENF signals given different counter forensic scenarios [29] [30]. On the other hand, the proposed AR model could provide a tool for an adversary to forge a multimedia signal using the anti-forensics operations discussed in [29] [30]. The model for the ENF signal can also apply to the generation mechanism of ENF signals, which can potentially be used to characterize the properties of a power grid and to estimate the grid-region in which a recording occurred. Such location estimation capabilities of the ENF signals using the proposed model are discussed in Chapter 5.

## 4.2 Autoregressive (AR) Model for ENF

An autoregressive model offers a common way of analyzing a correlated time series. An AR process is a regression of the current value of a time series based on

the previous observed values. A time-series  $u(n), u(n-1), \dots, u(n-M)$  represents realization of a real AR process of order  $M$ , denoted by  $\text{AR}(M)$ , if it satisfies the following difference equation:

$$u(n) + a_1u(n-1) + \dots + a_Mu(n-M) = v(n), \quad (4.1)$$

where  $a_1, a_2, \dots, a_M$  are the constants representing a stable filter, and known as the AR coefficients, and  $v(n)$  is a white noise process and uncorrelated with  $u(n-1), u(n-2), \dots, u(n-M)$ . The process  $v(n)$  brings randomness to  $u(n)$  and is known as an *innovation* process. In terms of linear filtering, an  $\text{AR}(M)$  can be generated by feeding  $v(n)$  as an input to an all pole filter with a  $z$ -transform given by  $A(z) = \frac{1}{1 - \sum_{m=1}^M a_m z^{-m}}$ . Additionally, if  $v(n)$  is a zero-mean Gaussian process of power  $\sigma_v^2$ , the output process  $u(n)$  is also a zero-mean Gaussian wide sense stationary (WSS) process, and its power spectral density is a function of the filter parameters  $a_1, a_2, \dots, a_M$ , and  $\sigma_v^2$ . Given such a process  $u(n)$ , an estimate of the model coefficients and statistics of  $v(n)$  can be obtained by solving a set of linear equations, known as the Yule-Walker equations. A detailed discussion on AR processes can be found in [31].

### 4.2.1 Statistics of ENF signals

Let  $f(n)$  denote an ENF signal at any time  $n$ , and let  $\mathbf{F}(n) = [f(n), f(n+1), \dots, f(n+N-1)]^T$  represent a vector of  $N$  consecutive values of ENF at a given time instant  $n$ . The mean function  $\hat{\mu}(n)$  and the mean-normalized autocorrelation function  $\hat{r}(n, n+k)$  of ENF signal from a frame of  $N$  samples are estimated as

follows:

$$\begin{aligned}\hat{\mu}(n) &= \frac{1}{N} \sum_{l=0}^{N-1} f(n+l) \\ \hat{r}(n, n+k) &= \frac{1}{N} \sum_{l=0}^{N-1} [f(n+l) - \hat{\mu}(n)][f(n+l+k) - \hat{\mu}(n+k)].\end{aligned}$$

We record a ENF signal in the United States from the power main supply and estimate its instantaneous frequency, using the weighted energy spectrogram method described in [11] [32]. We obtain an estimate of the instantaneous frequency for every segment of 16 seconds.

The plots of  $\hat{\mu}(n)$  and  $\hat{r}(n, n+10)$  for  $N = 32$  (equivalently 512 seconds) and for different values of  $n$  are shown in Fig. 4.1(a) and 4.1(b), respectively. From these figures, we observe that the mean function  $\hat{\mu}(n)$  is very close to 60Hz. We also observe that the autocorrelation function is approximately independent of  $n$  for  $k = 10$ . Similar plots are obtained for different values of  $k$ . These results indicate that the ENF signal exhibits characteristics approximating a WSS process. More generally, if accounting for small variations in the autocorrelation function for different values of  $n$ , ENF signal can be approximated as a piecewise-WSS in small segments. Furthermore, the probability density function (pdf) of  $f(n)$  follows a Gaussian distribution, as shown in Fig. 4.2(a), for a 25-hours long ENF signal recording from the power supply. Based on these observations, we model the ENF signal  $f(n)$  as a Gaussian process that is piecewise-WSS with a mean value 60 Hz. To make it a zero-mean process, we subtract the nominal value of 60 Hz from the ENF signal. In our subsequent discussions, the mean-normalized version of  $f(n)$  is assumed to be a zero-mean process.

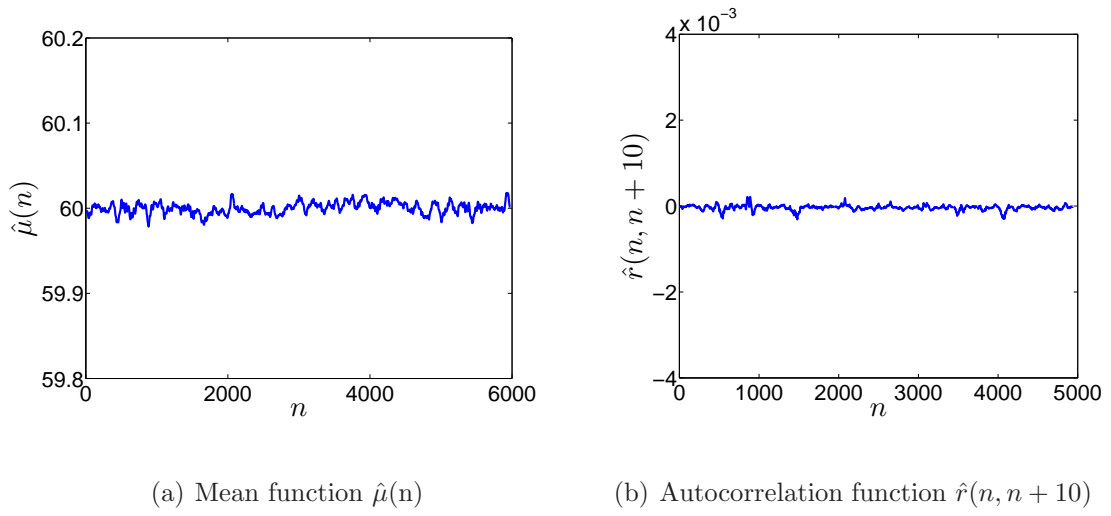


Figure 4.1: The mean and the autocorrelation function of a ENF signal recording.

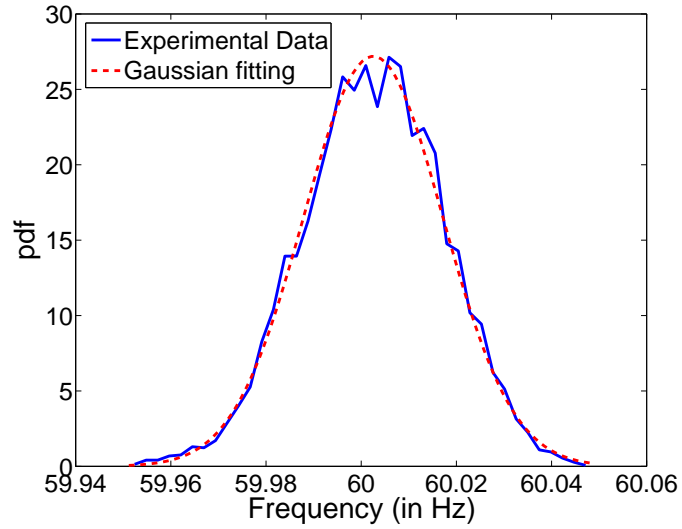


Figure 4.2: pdf of  $f(n)$  for a 25-hours long power-ENF recording.



## 4.2.2 Box-Jenkins Test for Model Validation

In the previous subsection, we observed that the ENF signal follows the properties of a Gaussian piecewise-WSS signal. In this subsection, we propose to use an AR process to characterize ENF signals. For this purpose, we use the Box-Jenkins methodology, which is widely used to model the time-series signals in statistics [27]. Box-Jenkins methodology applies to the signals following stationary properties. For non-stationary signals, the series can be converted first to a stationary series by pre-processing operations such as de-trending and de-seasoning [27].

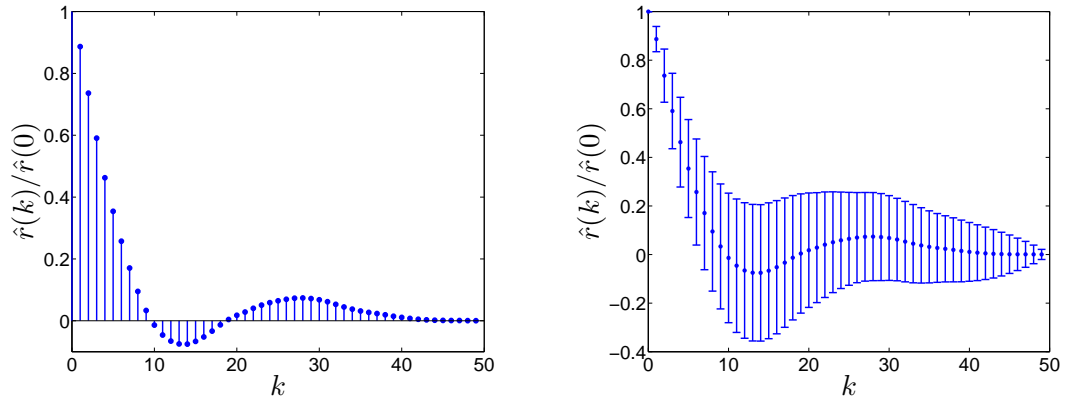
The Box-Jenkins method involves three steps to model a stationary series: 1) model identification, 2) parameter estimation, and 3) diagnostic testing. If the results from Step 3 satisfy the hypothesized model, we adopt the hypothesize model as the underlying model behind the observed process. The three steps are described below in detail.

**1. Model Identification:** The first step in model identification is to determine the stationarity of the signals. As discussed in Section 4.2.1, ENF signals exhibit WSS properties. Alternatively, the autocorrelation function can also be used to detect stationarity and hypothesize the model selection. The autocorrelation plot of the signal provides this information. We use the observations from Table 4.1 to determine the underlying signal model.

A plot of sample autocorrelation function for the ENF signal extracted from the US East grid is shown in Fig. 4.3(a). From this figure, we observe that the autocorrelation function decays exponentially to zero. Since it is only one instant of

Table 4.1: Hypothesis model of Box-Jenkins methodology.

Shape of $r(k)$	implied model	order estimation
Exponential, decaying to zero	AR	Use PACF for order estimation
One or more spikes	MA	Order corresponds to $k$ when $r(k) = 0$
Decaying to zero after few lags	ARMA	Cannot be determined
No decay to zero	Not stationary	Use differencing to make stationary



(a) sample autocorrelation function,  $r(k)$       (b) distribution of autocorrelation function,  $r(k)$

Figure 4.3: Plot of autocorrelation function.

the autocorrelation function, we plot the mean value of autocorrelation value at a lag  $k$ , and variance of that distribution is shown by bars for a database of 25-hours of ENF signal in Fig. 4.3(b). From this figure, we see that the decay of autocorrelation value can be approximated as exponential. Based on this observation, we model the ENF signal as an autoregressive process, as suggested by Table 4.1.

To determine the order  $p$  of the underlying  $AR(p)$  process, we use the partial autocorrelation function (PACF). Given the observations  $f(n)$  and  $f(n+k)$ , the

PACF denoted by  $\alpha(k)$  is defined as the correlation between  $f(n)$  and  $f(n+k)$ , given the intermediate observations  $f(n+1), f(n+2), \dots, f(n+k-1)$ . Mathematically, PACF  $\alpha(k)$  is defined as:

$$\alpha(k) = \frac{E[f(n), f(n+k)|f(n+1), \dots, f(n+k-1)]}{\sqrt{E[f(n)^2|f(n+1), \dots, f(n+k-1)]} \sqrt{E[f(n+k)^2|f(n+1), \dots, f(n+k-1)]}}. \quad (4.2)$$

For an AR( $p$ ) process, the value of  $\alpha(k) = 0$  for  $k > p$ , as  $f(n+p+1)$  is uncorrelated with  $f(n)$  for an AR( $p$ ) process if the intermediate observations  $f(n+1), f(n+2), \dots, f(n+p)$  are known. We use this property of  $\alpha(k)$  to determine the order of the AR model for ENF signals. A sample plot of  $\alpha(k)$  from a segment of US East Coast ENF recording is shown in Fig. 4.4. In this figure, the red line indicates the 95% significance level, indicating the level above which the value of  $\alpha(k)$  cannot be neglected. Since the ENF signal is modeled as a piecewise-WSS signal, we plot the distribution of  $\alpha(k)$  for the US East Coast ENF data in Fig. 4.5. From this figure, we see that for  $k > 2$ , the value of  $\alpha(k)$  mostly lies within the insignificance region. Based on this observation, we model the ENF signal as an AR process of order 2. Mathematically, the process can be written as following:

$$f(n) = a_1(n)f(n-1) + a_2(n)f(n-2) + v(n). \quad (4.3)$$

Since  $f(n)$  is a piecewise stationary process to take into account sudden glitches in the frequency, the value of  $a_1(n)$  and  $a_2(n)$  may differ in each segment.

After hypothesizing the underlying model, we go to Step 2 of the Box-Jenkins methodology to estimate the parameters of the underlying AR(2) process.

**2. Parameter Estimation:** From Step 1, we model the ENF signal as a piece-

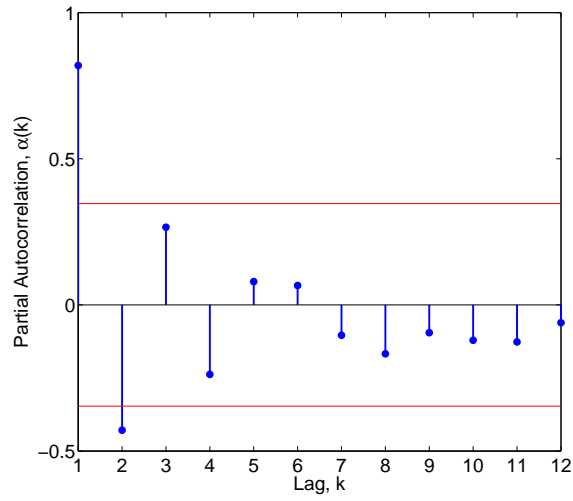


Figure 4.4: sample PACF  $\alpha(k)$ .

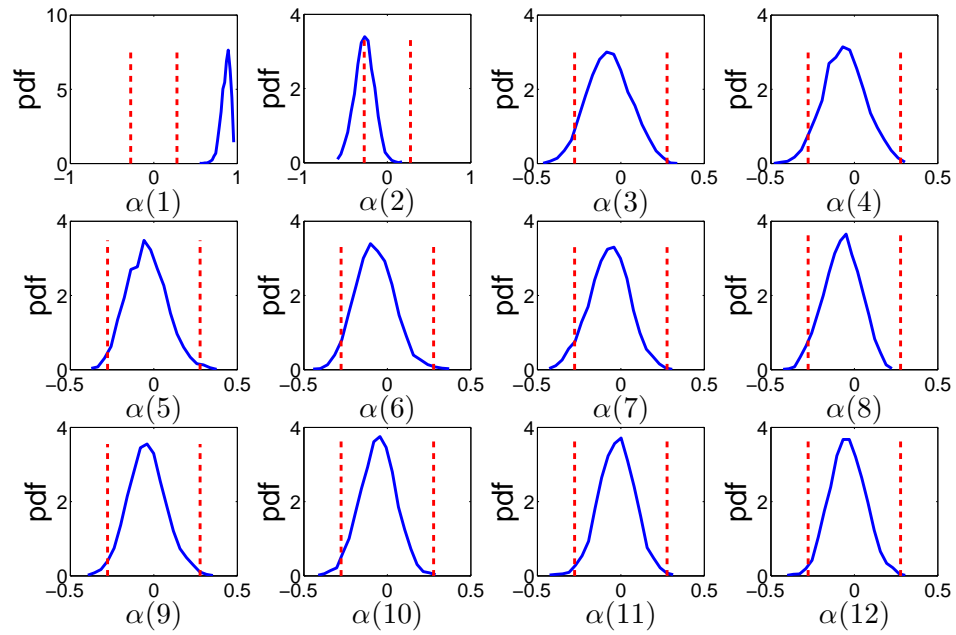


Figure 4.5: Distribution of partial autocorrelation function  $\alpha(k)$ . Red dashed lines represent 95% significance level.

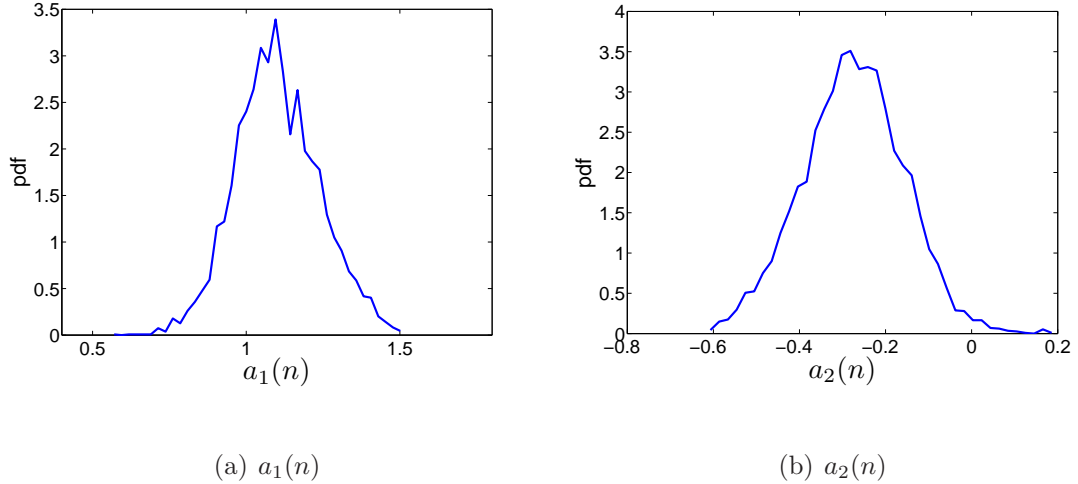


Figure 4.6: Distribution of AR(2) parameters estimated using the Yule-Walker equations.

wise AR(2) process. We estimate the AR filter coefficients  $a_1(n)$  and  $a_2(n)$  for each segment using the Yule-Walker equations. The plot of the distribution of estimated parameters  $a_1(n)$  and  $a_2(n)$  for a segment length of 50, with instantaneous frequency estimated for every 16-second segment, are shown in Fig. 4.6(a) and 4.6(b), respectively. From these figures, we observe that the values of  $a_1(n)$  and  $a_2(n)$  each follows a distribution similar to the Gaussian distribution. As the AR coefficients of the signal vary, we call the model as a time-varying AR model.

We also calculate the residue process  $v(n)$  by passing  $f(n)$  through a FIR filter, given by  $1 + a_1(n)z^{-1} + a_2(n)z^{-2}$ . The detailed discussion on Yule-Walker equations and obtaining the residue process can be found in [31].

**3. Diagnostic Testing:** This step is used to evaluate the fitness of the hypothesized model in Step 1, using the statistics of residual process  $v(n)$  obtained from Step 2. Diagnostic testing aims to evaluate if the residual process follows the property of

white noise. If this is the case, then the hypothesized model can be used to represent the given process. The whiteness of the residual process can be determined by checking its randomness at each time-lag. However, the Box-Jenkins methodology uses a more sophisticated statistical test to determine the whiteness of the residual process. This test is an adaptation of the chi-square test and looks to test the overall randomness of the residue rather than considering the randomness at each time-lag. The test involves evaluating the  $Q$ -statistics as following:

$$Q(m) = N(N + 2) \sum_{k=1}^M \frac{\hat{\rho}_k^2}{N - k}, m \ll N, \quad (4.4)$$

where  $\hat{\rho}_k$  is the correlation between the residue process  $v(n)$  and  $v(n - k)$  defined as following:

$$\hat{\rho}_k = \frac{\sum_{n=1}^{N-k} v(n)v(n - k)}{\sum_{n=1}^{N-k} v(n)^2}, \quad (4.5)$$

where  $N$  is the length of the segment. According to this test,  $Q(m)$  follows a chi-square distribution with  $m - p$  degrees of freedom, where  $p$  are the number of parameters in the model. For our case of AR(2) model, the value of  $p = 2$ ,

$$Q(m) \sim \chi_{m-2}^2. \quad (4.6)$$

The mean and the variance statistics of  $Q(m)$  are given by,

$$\begin{aligned} E [Q(m)] &= m - 2 \\ E [(Q(m) - E [Q(m)])^2] &= 2(m - 2). \end{aligned} \quad (4.7)$$

We plot the mean and the variance of  $Q(m)$  for the residue process obtained after passing the ENF signal through an AR filter of order 2 with coefficients obtained using the Yule-Walker equations. This plot is shown in Fig. 4.7. From this

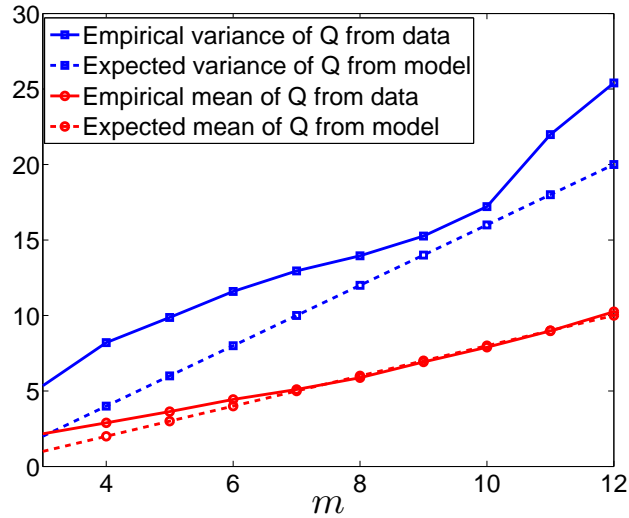
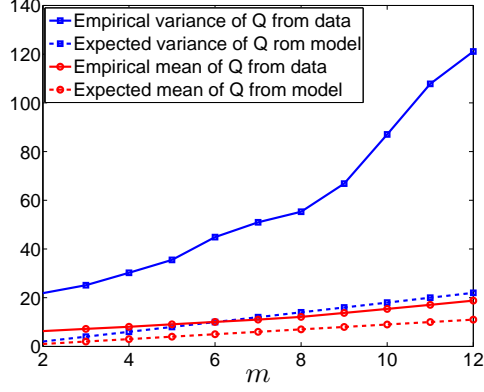


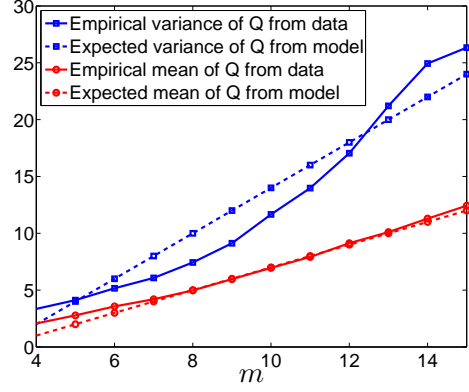
Figure 4.7: Mean and variance of  $Q(m)$  for the residue process  $v(n)$  obtained by assuming an AR(2) process for ENF signal.

figure, we observe that the expected value of the mean, and the variance of the  $Q$ -statistics closely follows the empirical values obtained from the  $Q$ -statistics of residue process. We also plot the statistic of  $Q(m)$  for residue process  $v(n)$  obtained under the assumption that our data follows an AR(1) and AR(3) model. These plots are shown in Fig. 4.8(a) and 4.8(b), respectively. From these plots, we observe a large mismatch between the expected statistics and the empirical statistics of  $Q(m)$  for AR(1) model. For AR(3) and other higher order models, these statistics remain close to the expected statistics. However, a higher model may lead to an overfitting of the data. So, we use the alternative model order selection criterion such as Bayesian Information Criterion (BIC) and the Akaike information criterion (AIC), which include a penalty term for the higher orders to avoid overfitting [33] [34].

AIC and BIC provides a quantifiable way of model order selection. The basic



(a)  $Q(m)$  test for AR(1) assumption



(b)  $Q(m)$  test for AR(3) assumption.

Figure 4.8: Mean and variance of  $Q(m)$  for the residue process  $v(n)$  obtained by assuming different orders of AR process for ENF signal.

idea behind these criteria incorporates is that there is a trade-off between the exactness of fit and the model complexity. By increasing the model order, data can fit to a model with increased likelihood. To avoid overfitting, AIC and BIC introduce a penalty term for the number of parameters in the model. For our case, the objective functions to minimize for the AIC and the BIC are given by the following equations:

$$AIC(k) = 2k + N \cdot \ln \left( \frac{\sum_{i=1}^N v_i^2}{N} \right) \quad (4.8)$$

$$BIC(k) = k \cdot \ln(N) + N \cdot \ln \left( \frac{\sum_{i=1}^N v_i^2}{N} \right) \quad (4.9)$$

The plots of  $AIC(k)$  and  $BIC(k)$  for different model orders  $k$  for  $N = 50$  are given in Fig. 4.9. From this figure, we observe that  $AIC(k)$  and  $BIC(k)$  reach their minimum value for  $k = 2$ . Based on these results, we conclude that ENF signals follow a time-varying AR(2) process.



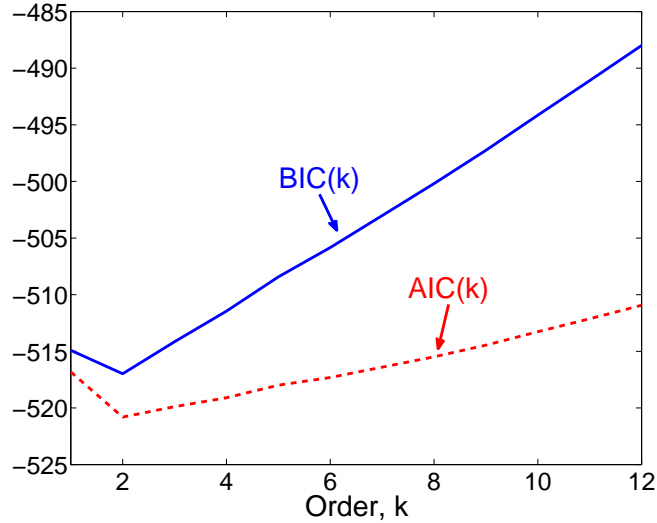


Figure 4.9: AIC(k) and BIC(k) for the residue process  $v(n)$ .

## 4.3 Timestamp Verification as Hypothesis

### Testing

In this section, we study the performance of the proposed ENF signal model for a timestamp verification application. In this application, each query multimedia file  $Z$  is assumed to contain an embedded timestamp in the metadata denoted by  $n$ , representing the time-of-recording claimed in the file. We want to verify the authenticity of this timestamp, i.e, to determine if the recording actually took place at time  $n$ . We use a hypothesis testing framework to study the performance of timestamp verification using ENF signal. We extract ENF signal from query  $Z$  using a bandpass filter, followed by instantaneous frequency estimation, and denote it by  $\underline{\mathbf{W}} = [w(0), w(1), \dots, w(N-1)]^T$ , where  $N$  is the length of the ENF signal  $\underline{\mathbf{W}}$

extracted from the file.

### 4.3.1 Matching using ENF sequences

We start with a highly simplified model to help gain insights on the ENF matching problem while retaining analytical tractability. Let us denote an  $N$ -point reference ENF signal at time instant  $n$  as  $\underline{\mathbf{F}}(n) = [f(n), f(n+1), \dots, f(n+N-1)]^T$ , which is stored in a database available to the detector. This database serves as a reference for the timestamp verification. Such an ENF database can be built by a continuous recording of the voltage signal from the power main supply and extracting the ENF signal using the instantaneous frequency estimation methods, as described in [32] [35]. Alternatively, the database can also be obtained directly from the power distribution companies as they generally keep a record of ENF signal [4].

As the sensitivity of various multimedia recording devices may vary, and interfering signals may exist in the frequency band around the nominal ENF value, distortion may be introduced in ENF signal embedded in query  $Z$ . Based on such observations,  $\underline{\mathbf{W}}$  can be assumed to be a distorted version of  $\underline{\mathbf{F}}(\cdot)$  corresponding to the actual time-of-recording. We model this distortion using an additive white Gaussian noise vector  $\underline{\mathbf{C}}(n) = [c(n), c(n+1), \dots, c(n+N-1)]^T$  with distribution  $\mathcal{N}(\underline{\mathbf{0}}, \sigma_c^2 \mathbf{I})$ , where  $\mathbf{I}$  denotes a  $N \times N$  identity matrix.

Under the settings described above, we model the ENF signal based timestamp verification as a binary hypothesis testing problem. We define two hypotheses,  $H_0$

and  $H_1$ , as follows:

$$H_0 : \underline{\mathbf{W}} = \underline{\mathbf{G}}(n) + \underline{\mathbf{C}}(n),$$

$$H_1 : \underline{\mathbf{W}} = \underline{\mathbf{F}}(n) + \underline{\mathbf{C}}(n).$$

Under the null hypothesis, the ENF signal  $\underline{\mathbf{W}}$  is a sample from  $\underline{\mathbf{G}}(n)$ , which has the distribution of a multivariate AR(2) process. Since each dimension of vector  $\underline{\mathbf{G}}(n)$  is a correlated random variable, it is difficult to derive its exact statistics. However, the first and the second order statistics can be derived as follows:

$$\begin{aligned} E(\underline{\mathbf{G}}(n)) &= \underline{\mathbf{0}}_{N \times 1}, \\ E(\underline{\mathbf{G}}(n)\underline{\mathbf{G}}(n)^T) &= \mathbf{R}_{N \times N}, \end{aligned} \quad (4.10)$$

where  $\mathbf{R}_{N \times N}$  is the correlation matrix with its  $(i, j)$ -th entry given by  $r(i - j)$ , with  $r(\cdot)$  being the autocorrelation function of ENF signal  $f(n)$ . Given the vector  $\underline{\mathbf{F}}(n)$  corresponding to timestamp  $n$  attached with the query file, and the vector  $\underline{\mathbf{W}}$  corresponding to the ENF signal extracted from the query file, a similarity based detector is used to measure the similarity as follows:

$$S = \frac{\underline{\mathbf{W}}^T \underline{\mathbf{F}}(n)}{\sqrt{\sum_{i=0}^{N-1} w(i)^2} \sqrt{\sum_{i=0}^{N-1} f(n+i)^2}}. \quad (4.11)$$

Under the described framework, the detector decides the authenticity of an intrinsically embedded timestamp in the given query  $Z$  by comparing the value of  $S$  with a pre-defined constant  $\tau$ :

$$\delta_D(S) = \begin{cases} 1 : & H_1 \text{ is declared if } S > \tau, \\ 0 : & H_0 \text{ is declared if } S \leq \tau. \end{cases} \quad (4.12)$$

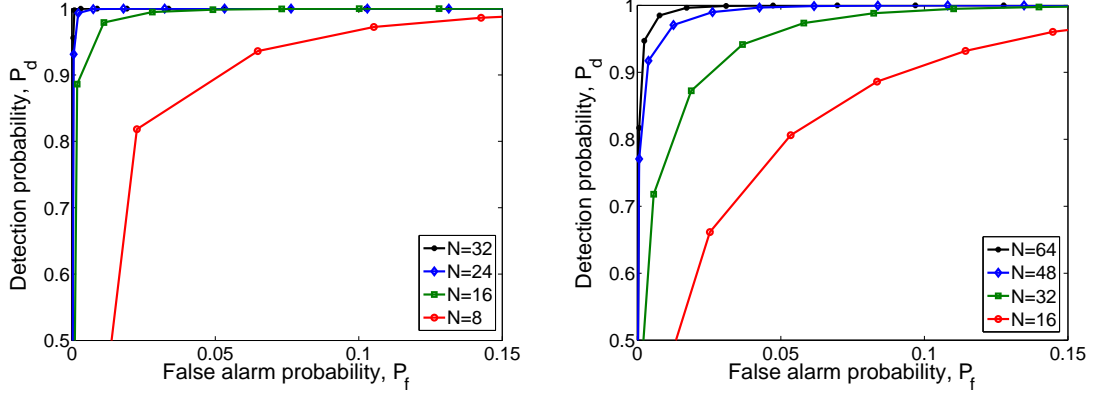
The performance of the timestamp verification under our model can be evaluated using the false alarm probability,  $P_f$ , and the detection probability,  $P_d$ , defined as:

$$\begin{aligned} P_f &= Pr(\delta_D = 1|H_0), \\ P_d &= Pr(\delta_D = 1|H_1), \end{aligned} \tag{4.13}$$

where  $Pr(\cdot)$  denotes the probability of the given event. The value of  $\tau$  presents a trade-off between  $P_f$  and  $P_d$ . For a practical system, the value of  $\tau$  should be chosen so that  $P_f$  is low and  $P_d$  is high. Due to a complex distribution of  $\mathbf{W}$ , it is difficult to derive the closed form expressions for  $P_f$  and  $P_d$ . To compare the performance of the proposed model with that of real data obtained from audio and power recordings, we use the Monte-Carlo simulations.

### 4.3.2 Results & Discussions

To obtain the Receiver Operating Characteristics (ROC) of the  $S$ -statistics defined in Eq. (4.11), we conduct Monte-Carlo simulations. We use the values obtained from an audio-power ENF database as the parameter values in our simulations. To build the audio-power ENF dataset, we record a 25-hour long audio signal using a microphone at a sampling rate of 1KHz and divide the recorded signal into different segment lengths. The beginning time index is stored in the metadata as the time-of-recording for each segment. Instantaneous frequency is estimated for a frame duration of  $T_{frame}$  seconds using the weighted energy spectrogram method described in Chapter 2. Concurrent power recordings are also conducted using a current sensing circuitry, as discussed in Chapter 2, and the weighted energy spec-

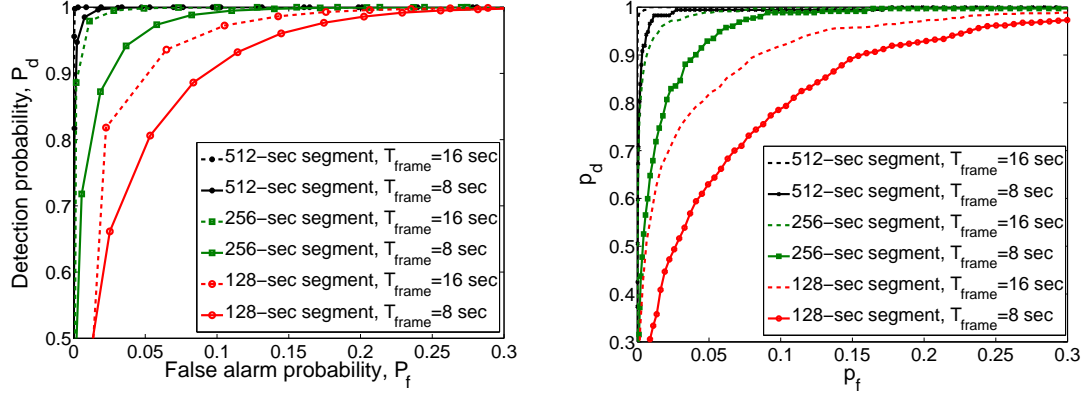


(a) SNR=15 dB (for  $T_{frame} = 16$  seconds)      (b) SNR=9.5 dB (for  $T_{frame} = 8$  seconds)

Figure 4.10: Receiver Operating Characteristics (ROC) of hypothesis detection framework for timestamp verification under the proposed AR(2) model for ENF signals.

rogram is used to estimate the ENF signal from this recording. The ENF estimates obtained from the power signal can be considered its cleanest, most readily available form, so an estimate of the signal-to-noise ratio SNR for audio-ENF signal may be obtained by subtracting the audio-ENF signal from the power-ENF signal to estimate the noise power. The instantaneous frequency estimation provides less robust estimates of frequency because the value of  $T_{frame}$  is decreased, so the SNR can be expected to decrease with the value of  $T_{frame}$ .

We plot the ROC of the  $S$ -statistics obtained using the Monte-Carlo simulations for different values of  $N$  for the SNR of 15 dB and 9 dB in Fig. 4.10(a) and 4.10(b), respectively. These SNR values correspond to the SNR of the audio-power ENF dataset for  $T_{frame} = 16$  seconds and  $T_{frame} = 8$  seconds, respectively. From these figures, we observe that the ROC characteristics improve as the value



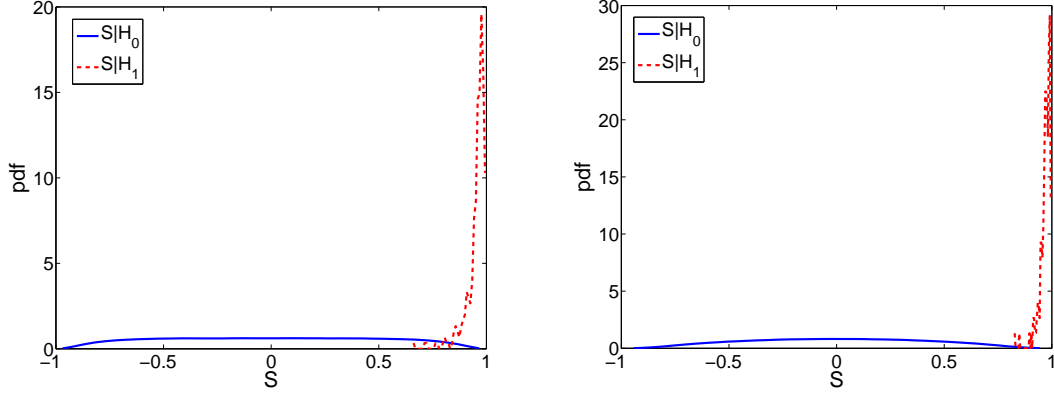
(a) Analytical results on proposed model

(b) Detection results on audio data

Figure 4.11: Comparison of the ROCs of the hypothesis detection framework for  $T_{frame} = 8$  seconds and  $T_{frame} = 16$  seconds for different query segment duration.

of  $N$  increases. For SNR corresponding to  $T_{frame} = 16$  seconds,  $P_d$  approaches 1 for  $P_f$  close to zero for  $N=32$  (equivalently 512 seconds long query segments). While for SNR corresponding to  $T_{frame} = 8$  seconds,  $P_d \approx 99\%$  for  $P_f = 1\%$ , when  $N=64$  (equivalently 512 seconds long query segments). This result indicates that for a given query duration, a higher value of  $T_{frame}$  works best to determine the matching between the multimedia and the power recordings using ENF signals. The relative performance of the detector for the values of  $N$  such that the query duration is the same for  $T_{frame} = 8$  seconds and  $T_{frame} = 16$  seconds is shown in Fig. 4.11(a). From this figure, it can be clearly seen that for the same query duration, using a higher value of  $T_{frame}$  improves the detection performance.

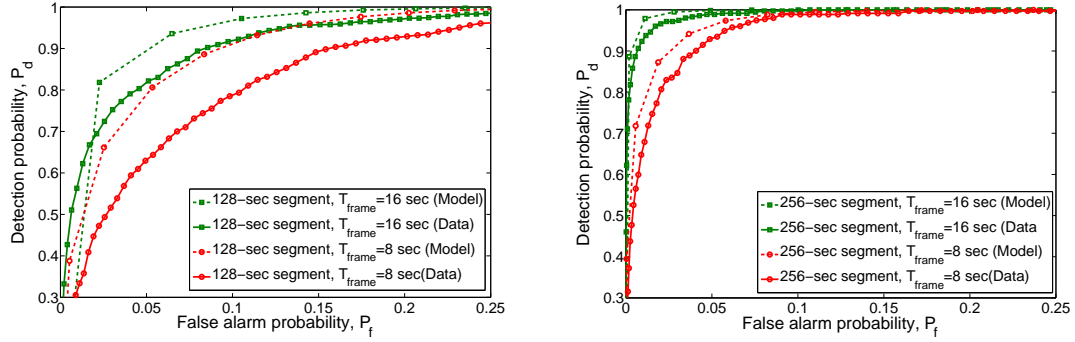
We also evaluate the performance of the proposed detector on timestamp verification for the power-audio ENF dataset described previously. Each audio file is given to the detector with every possible timestamp  $m \neq n$  for the false matching



(a) Segment length=256 seconds (i.e.,  $N=16$ ) (b) Segment length=512 seconds (i.e.,  $N=32$ )

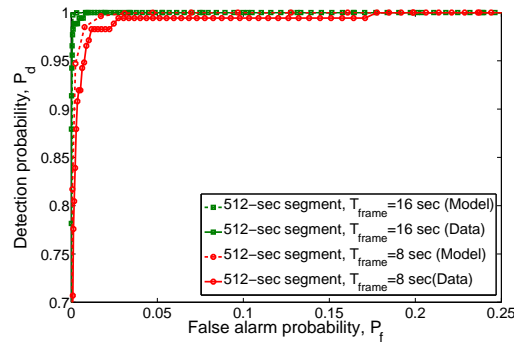
Figure 4.12: Pdf of  $S$  for the audio-power data for  $T_{frame} = 16$  seconds under  $H_0$  and  $H_1$  hypotheses for different query durations.

cases, and  $m = n$  for the correct matching case. This setting gives us 700, 350, and 175 correct matching samples for segment lengths of 128 seconds, 256 seconds, and 512 seconds, respectively in our database. The plots of the pdf of  $S$  obtained using this dataset under  $H_0$  and  $H_1$  hypotheses for  $T_{frame} = 16$  seconds are shown in Fig. 4.15(a) and 4.15(b) for  $N = 16$  and  $N = 32$ , respectively. We observe from these figures that the distribution of  $S$  spreads more uniformly in  $[-1,1]$  range under  $H_0$  for  $N = 16$  as compared with that of  $N = 32$ . As the value of  $N$  increases, the distribution of  $S$  under  $H_0$  closely follows a Gaussian distribution. We also notice that the distribution of  $S$  under  $H_1$  is concentrated around 1. The distribution of  $S$  under the two hypotheses partially overlap, which will cause errors in detection. The plots of the ROC for a query duration of 128 seconds, 256 seconds, and 512 seconds for  $T_{frame} = 8$  seconds and  $T_{frame} = 16$  seconds are shown in Fig. 4.11(b). From this figure, we observe similar trends as predicted from the ROC of the  $S$ -



(a) segment length=128 seconds

(b) segment length=256 seconds



(c) segment length=512 seconds

Figure 4.13: Comparison of the ROCs of the hypothesis detection framework for timestamp verification under the proposed model and the empirical data for different query duration.

statistics under the proposed model, as shown in Fig. 4.11(a). The performance improves with an increase in the value of  $T_{frame}$  for a fixed query duration, and for an increase in query duration for a fixed  $T_{frame}$ .

In Fig. 4.13(a)-(c), we compare the ROC performance of the hypothesis detection framework for timestamp verification under the proposed model and the empirical data for different query duration of 128 seconds, 256 seconds, and 512



seconds, respectively. From these figures, we observe that the ROCs obtained from the model closely match with that obtained from the empirical data for  $T_{frame} = 8$  seconds and 16 seconds. We also discover that the gap between the ROCs from the model and that of the empirical ENF data reduces as the query duration increases. These results indicate that best results are obtained using a large query segment for matching, and for a given segment length, a larger value of  $T_{frame}$  will provide better matching performance. Next, we study an improved method to compare the ENF signals from multimedia with that of power-recordings using the proposed model for ENF signals.

## 4.4 Improved Method for Matching ENF Signals

In the previous section, we have formulated the problem of ENF signal based timestamp verification as a binary hypothesis testing problem. We have observed that the distribution of the detection statistics  $S$  under  $H_0$  and  $H_1$  shows a significant amount of overlap. A major contributor to this overlap is the correlation within an ENF signal over time, as evidenced from the autoregressive model of ENF signal described in Section 4.2. Such correlation may lead to local peaks in the value of  $S$  at time shifts other than the true match and, in turn, a high false alarm probability. To improve the detection performance, we propose to use samples from the *innovation* process  $\underline{\mathbf{V}}(n) = [v(n), v(n+1), \dots, v(n+N-1)]^T$  for matching. Using the ENF signal model from Section 4.2, the process  $\underline{\mathbf{V}}(n)$  can be obtained after decorrelating the ENF signal  $\underline{\mathbf{F}}(n)$  at time  $n$  by filtering it through a filter

$H(z) = \frac{1}{A(z)} = 1 - a_1(n)z^{-1} - a_2(n)z^{-2}$ . The filter coefficient  $a_1(n)$  and  $a_2(n)$  can be estimated using the Yule-Walker equations applied to the ENF data  $\underline{\mathbf{F}}(n)$  at time  $n$ . Since  $v(n)$  is an i.i.d. sequence for an AR process, this methodology addresses the false alarm probability described above and may provide an improvement in the detection performance.

By passing  $\underline{\mathbf{W}}$  through the estimated filter  $H(z)$ , we obtain the corresponding query innovation process and denote it by  $\underline{\mathbf{W}}_d = [w_d(0), w_d(1), \dots, w_d(N-1)]^T$ . Under this setting, for an embedded timestamp  $n$  in the given query  $Z$ , we use  $\underline{\mathbf{W}}_d$  and  $\underline{\mathbf{V}}(n)$  for hypothesis testing. The two hypotheses now become:

$$H_0 : \underline{\mathbf{W}}_d = \underline{\mathbf{U}}(n) + \underline{\mathbf{D}}(n),$$

$$H_1 : \underline{\mathbf{W}}_d = \underline{\mathbf{V}}(n) + \underline{\mathbf{D}}(n),$$

where  $\underline{\mathbf{D}}(n) = [d(n), d(n+1), \dots, d(n+N-1)]^T$  is a vector of zero-mean colored Gaussian noise process, with its components given by  $d(n) = c(n) - a_1(n)c(n-1) - a_2(n)c(n-2)$ , where  $a_1(n)$  and  $a_2(n)$  are the AR coefficients for the corresponding segment. The power of the noise process  $d(n)$  is denoted by  $\sigma_d^2$ . Under the null hypothesis,  $\underline{\mathbf{W}}_d$  is a sample from  $\underline{\mathbf{U}}(n)$ . The distribution of the innovation sequences from ENF database,  $\underline{\mathbf{U}}(n)$ , can be modeled as  $\mathcal{N}(\underline{\mathbf{0}}, \sigma_v^2 \mathbf{I})$ . To measure the similarity between  $\underline{\mathbf{W}}_d$  and  $\underline{\mathbf{V}}(n)$ , we define a similarity based metric similar to Eq. (4.11) as follows:

$$\begin{aligned} S' &= \frac{\underline{\mathbf{W}}_d^T \underline{\mathbf{V}}(n)}{\sqrt{\sum_{i=0}^{N-1} w_d(i)^2} \sqrt{\sum_{i=0}^{N-1} v(n+i)^2}} \\ &\simeq \frac{1}{N} \frac{\underline{\mathbf{W}}_d^T \underline{\mathbf{V}}(n)}{\sqrt{\sigma_{wd}^2} \sqrt{\sigma_v^2}}, \end{aligned} \quad (4.14)$$

where  $\sigma_{wd}^2$  and  $\sigma_v^2$  are the variances of components in  $\underline{\mathbf{W}}_d$  and  $\underline{\mathbf{V}}(n)$ , respectively. In practice, the value of  $S'$  is obtained by using estimated values of  $\sigma_{wd}^2$  and  $\sigma_v^2$  in the denominator of Eq. (4.14). The estimated values of  $\sigma_{wd}^2$  and  $\sigma_v^2$  is computed as  $\hat{\sigma}_{wd}^2 = \frac{1}{N} \sum_{k=0}^{N-1} w_d^2(k)$  and  $\hat{\sigma}_v^2 = \frac{1}{N} \sum_{k=0}^{N-1} v^2(n+k)$ , respectively. The covariance matrix  $\mathbf{R}_d$  of the process  $\underline{\mathbf{D}}(n)$  is not diagonal because the noise is colored after passing through filter  $H(z)$ .  $(i, j)^{th}$  entry of  $\mathbf{R}_d$  is given by  $r_d(i-j)$ , and  $r_d(0) = \sigma_d^2$  and  $r_d(k) = 0 \forall |k| > 2$ . The distribution of the detection statistics  $S'$  under  $H_0$  and  $H_1$  can be written as follows:

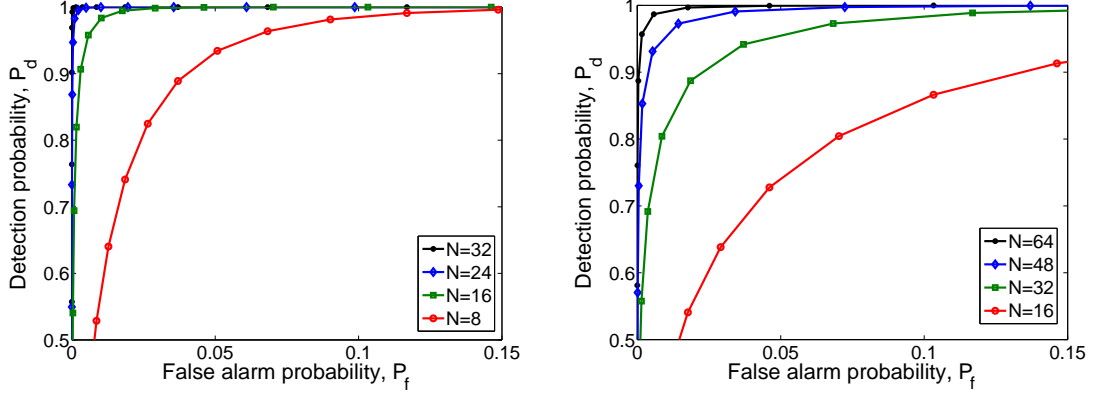
$$S'|H_0 \sim \mathcal{N}\left(0, \frac{1}{N}\right), \quad (4.15)$$

$$S'|H_1 \sim \mathcal{N}\left(\frac{1}{\sqrt{1 + \frac{1}{SNR'}}}, \frac{1}{N} \left(\frac{1}{1 + SNR'}\right)\right), \quad (4.16)$$

where  $SNR'$  is defined as  $\frac{\sigma_v^2}{\sigma_d^2}$ . The detailed proof of these expressions are provided in the Appendix at the end of this chapter. Note that the value of  $SNR'$  is significantly less than that of  $SNR$ , as filtering  $\underline{\mathbf{W}}$  using the filter  $1 - a_1(n)z^{-1} - a_2(n)z^{-2}$  leads to an increase in the noise power  $\sigma_d^2$  of the resulting signal  $\underline{\mathbf{W}}_d$ . As will be discussed in Section 4.4.1, even given this reduced signal-to-noise ratio, the detection performance of timestamp verification improves with matching using *innovation* sequences.

#### 4.4.1 Results on Audio

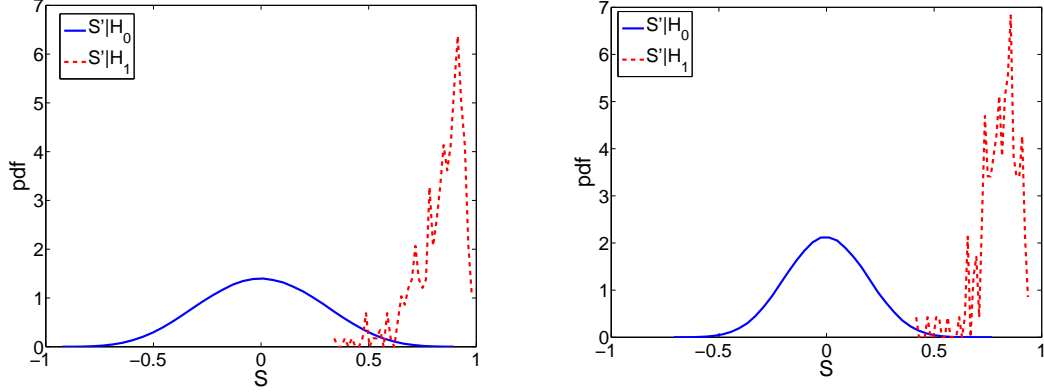
The ROC of the  $S'$ -statistics for different values of query length  $N$  are shown in Fig. 4.14(a) and 4.14(b) for  $SNR'=4.95\text{dB}$  and  $SNR'=-3.68\text{dB}$ . These values of  $SNR'$  correspond to the empirical values obtained from the dataset after deorrelating



(a)  $SNR' = 4.95$  dB (for  $T_{frame} = 16$  seconds)    (b)  $SNR' = -3.68$  dB (for  $T_{frame} = 8$  seconds)

Figure 4.14: Receiver Operating Characteristics(ROC) of the proposed AR decorrelation based hypothesis detection framework for timestamp verification using the model.

the power and audio ENFs by using an AR(2) decorrelation on both the ENFs estimated using  $T_{frame} = 16$  seconds and  $T_{frame} = 8$  seconds, respectively. Firstly, these empirical values of  $SNR'$  have values approximately 10dB less than those of  $SNR$  for the same  $T_{frame}$  size, when the ENF signals are matched directly. As the ROC performance is dependent of the noise level in the signal used for matching, the effect of decorrelation on ROC may be nullified by a decrease in SNR conditions, when matching is performed using innovation sequences. The trends in the ROC have similar results to the ROC obtained when matching occurred directly using the ENF signals. The ROC performance improves as the value of  $N$  improves because using more number of points for similarity measure is robust against noise. We also observe that for a fixed query duration,  $T_{frame} = 16$  seconds provides a better performance compared to  $T_{frame} = 8$ -seconds.



(a) Segment length=256 seconds

(b) Segment length=512 seconds

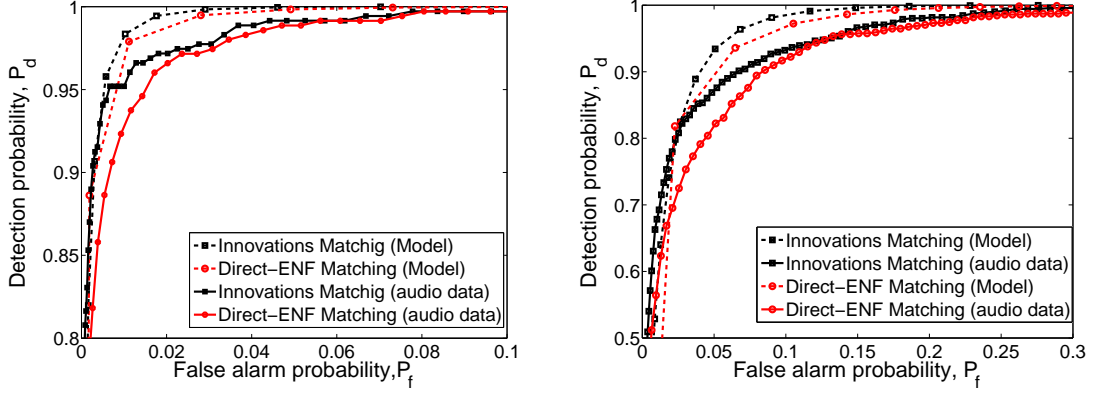
Figure 4.15: Pdf of  $S'$  for the audio-power data for  $T_{frame} = 16$  seconds under  $H_0$  and  $H_1$  hypotheses for different query durations.

For our audio experimental data, we use an AR(2) model to estimate the filter parameters  $a_1(n)$  and  $a_2(n)$  of  $A(z)$  for the power-ENF signal  $\underline{\mathbf{F}}(n)$  at time  $n$  using the Yule-Walker equations. To obtain  $\underline{\mathbf{V}}(n)$  and  $\underline{\mathbf{W}}_d$ , we filter the power-ENF signal  $\underline{\mathbf{F}}(n)$  and the query audio-ENF signal  $\underline{\mathbf{W}}$  by passing them through  $H(z) = 1 - a_1(n)z^{-1} - a_2(n)z^{-2}$ , where  $a_1(n)$  and  $a_2(n)$  are the AR coefficients obtained from  $\underline{\mathbf{F}}(n)$ . Using the same settings as described in Section 4.3.2, we conduct a timestamp verification operation by a direct ENF sequence matching and decorrelated *innovation* sequence matching, respectively. Fig. 4.15(a) and 4.15(b) show the pdf plots of  $S'$  under  $H_0$  and  $H_1$  hypotheses for  $T_{frame} = 16$  seconds for  $N = 16$  and  $N = 32$ , respectively. We observe that the distribution of  $S'$  under  $H_0$  and  $H_1$  follows a Gaussian distribution, as indicated by our analysis used to derive Eq. (4.15). We also note that the distribution of  $S'$  under  $H_1$  concentrates around 1. The distribution of  $S'$  under the two hypotheses partially overlap and will cause

some errors in detection.

The ROC curves for the detectors described in Section 4.3.1 and 4.4, when 256-second and 512-second long queries are used for timestamp verification, are shown in Fig. 4.16(a) and Fig. 4.16(b). From these figures, we observe that the performance of the detector on audio data compares well to the performance predicted using our analytical model. We further observe that using the innovation sequences for matching performs slightly better than using the ENF sequences directly, especially for a low false alarm rate. For example, the probability of detection increases from 92% to 95% for a false alarm rate of 1% for 256-second segments, when innovation sequences are used for matching in our experiments. Similarly for 512-second segments, the probability of detection increases from 82% to 87% for a false alarm rate of 5% when innovation sequences are used for matching. We note a significant improvement in the detection performance when the query clip duration increases from 128 seconds to 256 seconds, for matching with direct ENF sequences and the innovation sequences.

In addition, from Fig. 4.16(a) and 4.16(b), we observe a slight mismatch between the performance of the analytical model and the experimental results. This mismatch may occur due to the simpler binary hypothesis detection framework used in this paper. In general, correlation between query ENF sequence and ENF database also appears high for time index near the actual time-of-recording. A composite hypothesis, taking such correlated structures near the time index corresponding to time-of-recording into account, would be a more desirable choice to model  $H_0$ . The ROC performance of the proposed model may match better with



(a) Segment length=256 seconds

(b) Segment length=128 seconds

Figure 4.16: ROC characteristics of the correlation detector for ENF matching v.s. innovations matching at two query segment length for  $T_{frame} = 16$  seconds.

that from the ENF data using such a composite hypotheses detection framework.

#### 4.4.2 Results on Video

We use the proposed AR(2) model to demonstrate a performance improvement for a timestamp verification setting considered similar to experiments on audio. Under this setting, we assume that each video recording contains a timestamp attached to it as a metadata. We use a database of 40 video-clips, each 10-minutes in duration and containing a timestamp. We also create multiple files and modify their metadata timestamps to other possible timestamps in the database. These two cases correspond to the following binary hypotheses:

$$\begin{cases} H_0 : & \text{The attached timestamp is fabricated;} \\ H_1 : & \text{The attached timestamp is genuine.} \end{cases} \quad (4.17)$$

Under this setting, we estimate the  $S$ -value between the ENF signal extracted

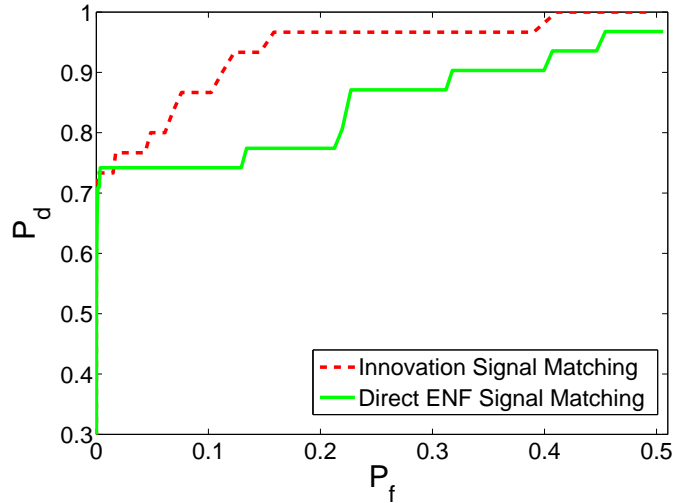


Figure 4.17: ROC for timestamp verification when using ENF signal and innovation sequences for matching.

from the video clip and the ENF signal extracted from the power signal corresponding to the timestamp attached to the recording. We then estimate the  $S'$ -value between the innovation signals obtained after decorrelating the ENF signals using the proposed AR(2) model. The attached timestamp is declared genuine if the values of similarity measure  $S$  and  $S'$  are greater than a threshold. From the ROC for this experiment shown in Fig. 4.17, we can observe an improvement in the detection performance when the innovation signals obtained after applying AR(2) model are used for matching, as compared to when direct ENF signals are used for matching. This happens because the side lobe that may be present in the normalized cross-correlation function (NCC) function of the ENF sequences given the correlated nature of the ENF sequences has been substantially mitigated in the NCC function of the innovation sequences.



## 4.5 Chapter Summary

In this chapter, we have proposed an analytical model for ENF signals based on autoregressive process. We have verified the model's fitness using the Box-Jenkins methodology. The proposed model has been used to study the problem of timestamp verification under a hypothesis detection framework. The trends in the receiver operating characteristics of the analytical model for different segment size used for matching are similar to those obtained from the experimental data. Based on the proposed model, a decorrelation based *innovation* process matching approach has been adopted to improve the performance of the timestamp verification under the proposed framework. The experimental results with audio data has demonstrated an improvement in the detection performance for a low value of false alarm rate. The binary hypothesis framework used in this chapter can be considered as a first-step exploration on the problem of ENF modeling for timestamp verification. Although the simplified hypothesis framework presented in this paper has a slight gap with the realistic characteristics as observed from our experiments, it shows that an AR process based decorrelation provides improvement in the detection performance, and this improvement is observed consistently when applied to audio data. The performance of the proposed model match more closely with the audio experimental data as segment length increases.

## Appendix: Derivation of Detection Statistics of $S'$

From Eq. (4.14),  $S'$  between  $\underline{\mathbf{W}}_{\mathbf{d}}$  and  $\underline{\mathbf{V}}(n)$  is defined as following:

$$S' = \frac{1}{N} \frac{\underline{\mathbf{W}}_{\mathbf{d}}^T \underline{\mathbf{V}}(n)}{\sqrt{\sigma_{wd}^2} \sqrt{\sigma_v^2}}, \quad (4.18)$$

where  $\sigma_{wd}^2 = \sigma_v^2 + \sigma_d^2$ . The distribution of  $\underline{\mathbf{W}}_{\mathbf{d}}$  under hypothesis  $H_0$  and  $H_1$  for a given  $\underline{\mathbf{V}}(n)$  is given by the following equations:

$$\underline{\mathbf{W}}_{\mathbf{d}}|H_0 \sim \mathcal{N}(\underline{\mathbf{0}}, (\sigma_v^2 \mathbf{I} + \mathbf{R}_{\mathbf{d}})) \quad (4.19)$$

$$\underline{\mathbf{W}}_{\mathbf{d}}|H_1 \sim \mathcal{N}(\underline{\mathbf{V}}(n), \mathbf{R}_{\mathbf{d}}) \quad (4.20)$$

where  $\mathbf{R}_{\mathbf{d}}$  is the correlation matrix of the noise  $\underline{\mathbf{D}}(n)$ . The noise process  $d(n)$  becomes a moving average (MA) process after  $c(n)$  is passed through  $H(z)$ , and the matrix  $\mathbf{R}_{\mathbf{d}}$  is not diagonal.  $(i, j)^{th}$  entry of  $\mathbf{R}_{\mathbf{d}}$  is given by  $r_d(i - j)$ , and  $r_d(0) = \sigma_d^2$  and  $r_d(k) = 0 \forall |k| > 2$ . As  $\underline{\mathbf{W}}_{\mathbf{d}}$  is a Gaussian multivariate distribution and  $S'$  is a linear function of  $\underline{\mathbf{W}}_{\mathbf{d}}$ , so  $S'$  also follows a Gaussian distribution. We derive the mean and the variance of  $S'$  as following:

$$\begin{aligned} E[S'|H_1] &= \frac{1}{N} \frac{E[\underline{\mathbf{W}}_{\mathbf{d}}^T|H_1]\underline{\mathbf{V}}(n)}{\sqrt{\sigma_v^2 + \sigma_d^2} \sqrt{\sigma_v^2}} \\ &= \frac{1}{N} \frac{\underline{\mathbf{V}}(n)^T \underline{\mathbf{V}}(n)}{\sqrt{\sigma_v^2 + \sigma_d^2} \sqrt{\sigma_v^2}} \\ &\stackrel{(a)}{\cong} \frac{\sigma_v^2}{\sqrt{\sigma_v^2 + \sigma_d^2} \sqrt{\sigma_v^2}} \\ &\cong \frac{1}{\sqrt{1 + \frac{1}{SNR'}}}, \end{aligned} \quad (4.21)$$

where  $SNR' = \frac{\sigma_v^2}{\sigma_d^2}$ .

$$\begin{aligned}
E[S'^2|H_1] &= \frac{1}{N^2} \frac{E[\mathbf{W}_d^T \mathbf{V}(n) \mathbf{V}(n)^T \mathbf{W}_d | H_1]}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&= \frac{1}{N^2} \frac{E[\text{trace}(\mathbf{W}_d \mathbf{W}_d^T \mathbf{V}(n) \mathbf{V}(n)^T) | H_1]}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&= \frac{1}{N^2} \frac{\text{trace}(E[\mathbf{W}_d \mathbf{W}_d^T | H_1] \mathbf{V}(n) \mathbf{V}(n)^T)}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)}. \\
\Rightarrow \text{Var}[S'|H_1] &= \frac{1}{N^2} \frac{\text{trace}(\mathbf{R}_d \mathbf{V}(n) \mathbf{V}(n)^T)}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&= \frac{1}{N^2} \frac{\sigma_d^2 \sum_{i=0}^{N-1} v_i^2 + 2r_d(1) \sum_{i=0}^{N-2} v_i v_{i+1} + 2r_d(2) \sum_{i=0}^{N-3} v_i v_{i+2}}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&\stackrel{(a)}{\cong} \frac{1}{N} \frac{\sigma_d^2 \sigma_v^2}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&= \frac{1}{N} \left( \frac{1}{1 + SNR'} \right), \tag{4.22}
\end{aligned}$$

where  $v_i$  is a notation used to represent  $v(n+i)$ , and (a) follows from the ergodicity assumptions  $\frac{1}{N} \sum_{i=0}^{N-1} v_i^2 \cong \sigma_v^2$ ,  $\frac{1}{N} \sum_{i=0}^{N-2} v_i v_{i+1} \cong 0$  and  $\frac{1}{N} \sum_{i=0}^{N-3} v_i v_{i+2} \cong 0$  because  $v(n)$  is a white noise process.  $SNR'$  is defined as the signal-to-noise ratio of audio innovations with respect to power innovations and  $SNR' = \frac{\sigma_v^2}{\sigma_d^2}$ . From Eq. (4.21) and (4.22),

$$S'|H_1 \sim \mathcal{N} \left( \frac{1}{\sqrt{1 + \frac{1}{SNR'}}}, \frac{1}{N} \left( \frac{1}{1 + SNR'} \right) \right). \tag{4.23}$$

The pdf for  $S'$  under  $H_0$  is derived using the similar steps:

$$E[S'|H_0] = \frac{1}{N} \frac{E[\mathbf{W}_d^T | H_0] \mathbf{V}(n)}{\sqrt{\sigma_v^2 + \sigma_d^2} \sqrt{\sigma_v^2}} = 0. \tag{4.24}$$

$$\begin{aligned}
\text{Var}[S'|H_0] &= \frac{1}{N^2} \frac{\text{trace}(E[\mathbf{W}_d \mathbf{W}_d^T | H_0] \mathbf{V}(n) \mathbf{V}(n)^T)}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&= \frac{1}{N^2} \frac{\text{trace}((\sigma_v^2 \mathbf{I} + \mathbf{R}_d) \mathbf{V}(n) \mathbf{V}(n)^T)}{(\sigma_v^2 + \sigma_d^2)(\sigma_v^2)} \\
&\stackrel{(a)}{\cong} \frac{1}{N} \tag{4.25}
\end{aligned}$$

where (a) follows from the similar steps used to derive Eq. (4.22). From Eq. (4.24) and (4.25), the distribution of  $S'$  under  $H_0$  can be written as follows:

$$S'|H_0 \sim \mathcal{N}\left(0, \frac{1}{N}\right). \quad (4.26)$$

# Chapter 5

## AR Model Parameters for Grid of Recording Classification

### 5.1 Chapter Introduction

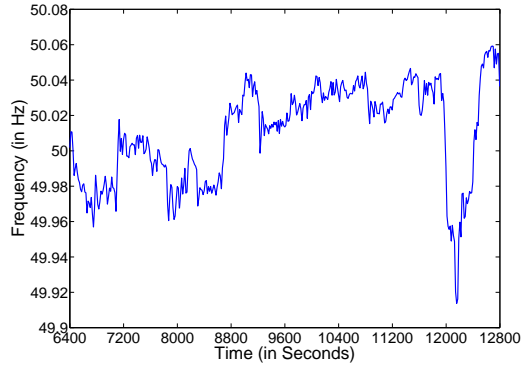
In the previous chapters, we discussed the timestamping property of ENF signals in such multimedia recordings as audio and video. ENF signal can be extracted from multimedia recordings that are influenced by electric power and compared with a power-ENF database to estimate or verify the time of recording. In this and the next chapter, we explore another important forensics problem of how to determine the location-of-recording of a given multimedia using power traces, which can be separated into two levels: a) region of recording as delineated by independent power grid, e.g., the US east, US west, Ireland, China, India, etc.; b) more specific area of the recording within a given region, e.g., to determine whether a recording happened in Maryland, North Carolina, or Massachusetts within the US east. A system that

possesses the capability of identifying a recording's location can be useful in many multimedia forensics and security applications as it allows identification of the origins of audio and videos originating from terrorist organizations, ransom demands, or child pornography and exploitation [36]. This sort of system can also be used for commercial applications such as automatic geo-tagging of user created multimedia data on social networking websites such as Twitter, YouTube, and Facebook.

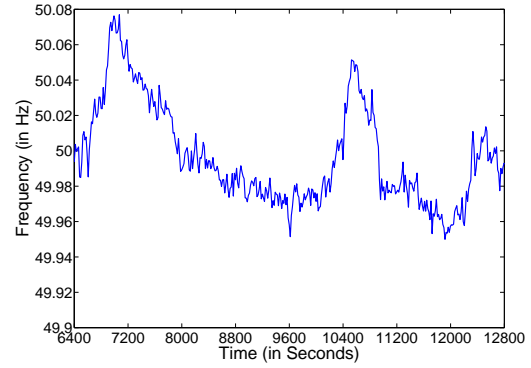
From our discussions in Chapter 2 and 3, we have established that the main trend of ENF signal variations at different areas of the same grid are the same at a given time due to the interconnected nature of the grid. Conversely, the ENF signal variations across different power grids are unlike, as the grids operate independently. In the presence of concurrent power recordings from the candidate grids, ENF signals from the given multimedia recording can be compared with the ENF database from the candidate grids to estimate the grid-of-recording. However, the record of concurrent power-ENF references from the candidate grids may not always be available. In this chapter, we focus on the problem of identifying the grid-region of a multimedia recording without having concurrent power references. In particular, we explore the grid characteristics from the AR modeling parameters, described in Chapter 4, as the features. We observe that some characteristics of ENF signals from different grids can be described using the AR modeling parameters. We build a support vector machine based classifier using these features to learn a classification model that can differentiate recordings made in various grids.

## 5.2 ENF Variations Across Different Grids

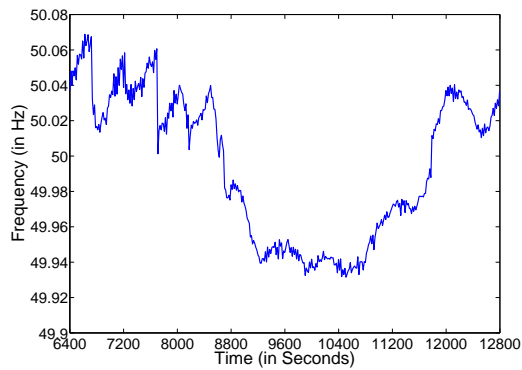
The statistical modeling presented in Chapter 4 reveals some highly interesting properties about the behavior and function of power grids. We have observed differences in the nature of ENF variations in different grids. We plot sample ENF signals for seven different grids in Figs. 5.1(a)- 5.1(d) and Figs. 5.2(a)- 5.2(c) for the European and the Asian grids of a Spanish island of Canary, Turkey, Ireland, and China, and the North American grids of US east, US west, and Quebec, respectively. The Spanish island, Turkey, Ireland, and China have a 50 Hz nominal ENF value, while US east, US west, and Quebec have a 60 Hz nominal ENF. From these figures, we observe the differences in the ENF signal variations on their particular grids. For example, the rate of change of ENF much higher in Quebec as compared with the other two grids on the continental North America. Similar observations can be made from the ENF plots of European and Asian grids. For example, ENF patterns for China tend to vary at a different rate than that of Ireland. ENF signal in Ireland shows a tendency to drift longer before returning to the nominal value, as compared with China and the Spanish island. Generally, the nature of fluctuations in the ENF depends on the size of the grid, i.e., a smaller capacity grid is less controlled and may exhibit a higher range of fluctuation, while a larger capacity grid tends to be more tightly controlled and may exhibit a smaller range of fluctuations [6]. For example, ENF signals in the Spanish island, Turkey, and Ireland power grids vary more greatly than the US east grid, as can be seen from Figs. 5.1(a)- 5.1(c)



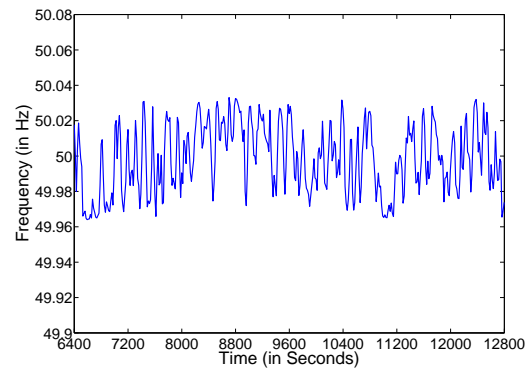
(a) Spanish island



(b) Turkey



(c) Ireland



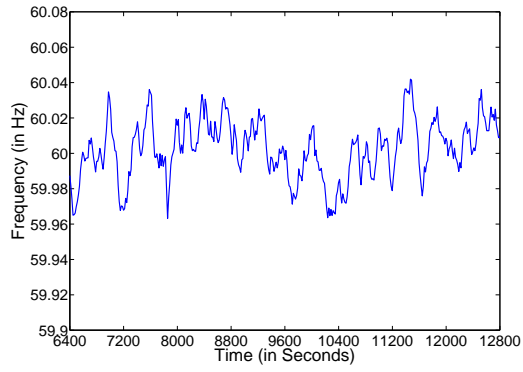
(d) China

Figure 5.1: Sample ENF plots for European and Asian grids.

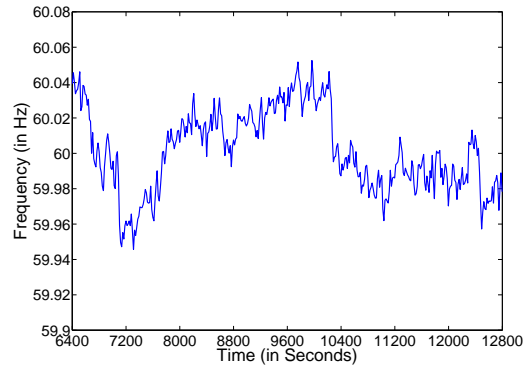
and 5.2(a).

The ENF variations in the power grids happen with a dynamically changes in load demand and supply. Power grids employ a frequency control mechanism to regulate the variations. When the frequency drops due to an increase in the load, the control mechanism senses the fluctuations and starts drawing power from adjoining areas/additional generators. This feedback compensates for the increased load to bring an increment in the supply frequency to bring it close to the nominal value. A similar feedback mechanism decreases the frequency value when supply

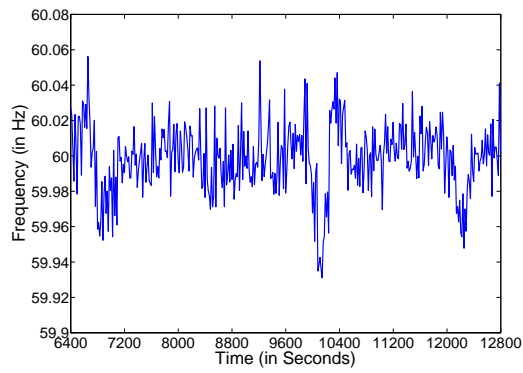




(a) US east



(b) US west



(c) Quebec

Figure 5.2: Sample ENF plots for continental North American grids.

frequency surges due to a drop in the load. Given the interconnected nature of the power grid, the ENF variations remain similar across a geographical region. Hence, the frequency control mechanism can also be considered similar for the entire power grid. As the current value of the ENF depends on its past values, the autoregressive model for ENF signals, described in Chapter 4, reflects such dynamics of the frequency control mechanism. An autoregressive model can be represented using autoregressive coefficients and noise power of the innovation process. We hypothesize that these modeling parameters can be used to characterize a grid, and can facilitate the identification of the grid-of-recording for any recording influenced by the ENF signals.

## 5.3 AR Modeling Parameters as Features

### 5.3.1 Feature Description

To understand the characteristics of the AR coefficients used to model the ENF signals in the power grids, we conduct multiple power recordings across different geographical regions covered by their own power grids. From each recording, we take a 8-minute long segment as one instance of recording and extract the ENF signal from it for a frame size of  $T_{frame} = 16$  seconds using the weighted spectrogram method described in Chapter 2. We then obtain a 30-points ENF signal for each 8-minute instance of recording in the same fashion. We extract the following features from this 30-point ENF signal:

*Average Frequency:* Two main nominal levels of ENF exist around the world, namely 50 Hz and 60 Hz, that can provide a basic 2-class information about the geographical location in terms of the power grid where the recording occurred. We compute the average frequency,  $F_{ave}$ , of the 30-point ENF signal obtained from the given recording example and use it as a feature for grid-region of location classification.

*AR Modeling Parameters:* We extract the AR(2) parameters,  $a_1$  and  $a_2$ , and the noise innovation power,  $\sigma_v^2$ , from the zero-mean normalization of ENF signal obtained after subtracting  $f_{ave}$  from the signal. Each recording instance of 8-minute long is represented by the following feature vector:

$$\underline{\mathbf{X}} = [f_{ave}, a_1, a_2, \sigma_v^2]^T \quad (5.1)$$

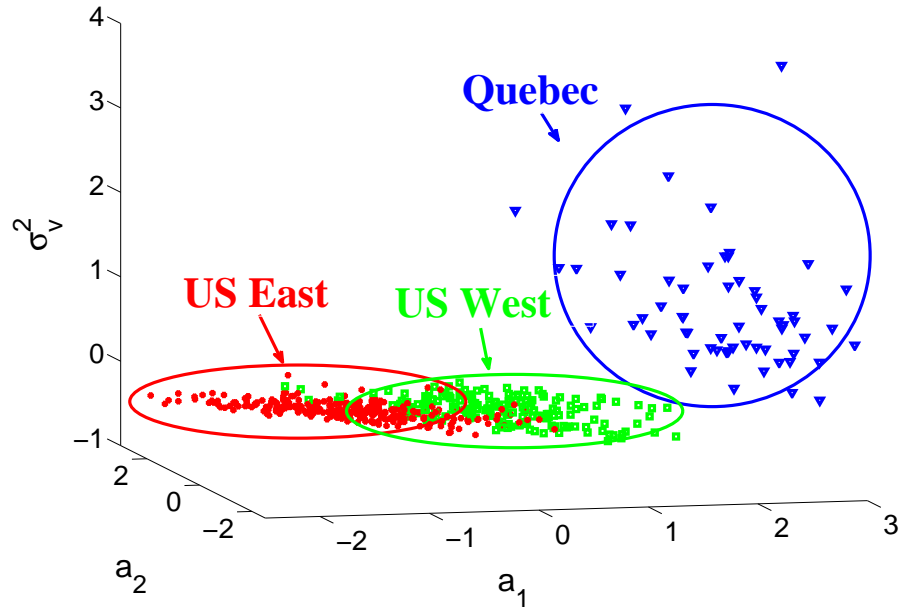
We obtain feature vectors for multiple instances of recordings from different power grids. Since the dynamic range of each feature within the vector can vary significantly, we normalize features using a standard normal procedure that makes each feature dimension have a zero-mean and unit-variance. This can be achieved by calculating the mean and the standard deviation along each feature dimension, then subtracting the mean followed by a division with the standard deviation from each feature value. A plot of the normalized AR modeling parameter features obtained from multiple instances of 8-minute signals across different power grids of Euro-Asia (50 Hz nominal ENF) and North America (60 Hz nominal) are shown in Fig. 5.3(a) and 5.3(b), respectively. From these figures we observe that the feature points obtained from the data from the same grid form clusters. For example, in North American grids, the features obtained from the power recordings obtained from the

Quebec grid form a non-overlapping cluster separate from the features obtained from the US east and the US west recordings. We also observe a partial overlap between the feature clusters from the two US grids, indicating that it may not be possible to classify the recordings from the US east and the US west with a high confidence by a linear classifier built on only AR parameters as features. Some overlap between clusters also occur for the European and the Asian grids. However, it may be possible to build a non-linear classifier, such as through the proper choices of kernels, to provide a better separation boundary.

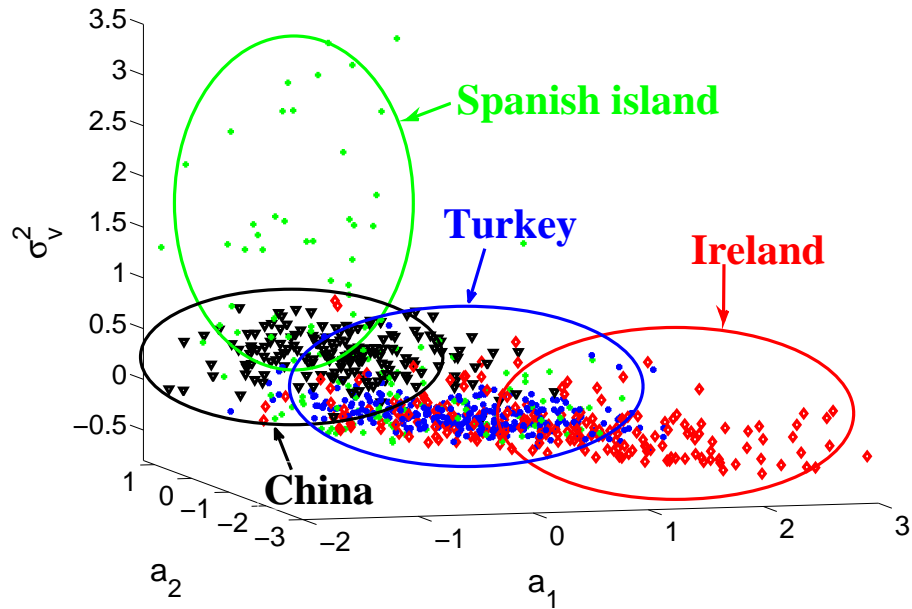
### 5.3.2 Experimental Setup

We build a multi-class classifier using a support vector machine (SVMs) method that has shown excellent classification accuracy in classifying the complex data in a number of applications [37]. We use the LibSVM implementation of a multi-class SVM for training a location classifier [38]. We have collected power and audio recordings from three grids in North America: the US east, the US west, and the Quebec interconnection; and from four grids in Europe and Asia: the Spanish island, Turkey, Ireland, and China. As discussed before, each training and testing examples contain features extracted from an 8-minute long recording.

The total numbers of training and testing examples from the power recordings across different grids in our database are available in Table 5.1. This table also shows the number of audio examples, which will be used for testing purposes for the classifier trained on features obtained from the power recordings. As the



(a) North American Grids: US east, US west, Quebec



(b) European and Asian Grids: Spanish island, Turkey, Ireland, China

Figure 5.3: AR parameters for location-of-recording classification for seven grid-regions.

training data for each class is unbalanced, i.e., the number of training examples are different for each class, we use a weighted-SVM implementation of the LibSVM. The weighted-SVM approach avoids over-fitting due to the unbalanced data by weighing the misclassification cost of each class as inversely proportional to the number of training examples present in that class. As non-linear classifiers demonstrate better classification accuracy as compared with the linear classifiers, we adopt a radial basis function (RBF) kernel to train the SVM classifier. RBF is a Gaussian-like shape non-linear kernel that has shown a superior classification results as compared to other kernels for many applications. The RBF kernel requires selection of two parameters, the cost  $C$  for each misclassification and the regularization parameter  $\gamma$  to control the spread of the RBF kernel. We use a 10-fold cross-validation strategy to search for the best values of these parameters. A contour plot for the validation accuracy for different values of  $C$  and  $\gamma$  is shown in Fig. 5.4. From this figure, we observe that the best validation accuracy of 78.10% for the training data is obtained for  $C = 20$ , and  $\gamma = 2^{-8}$ . We use these values of  $C$  and  $\gamma$  to train a multi-class SVM model.

### 5.3.3 Results and Discussions

We evaluate the classification accuracy of the proposed grid-region of recording classifier on the power and audio recordings. Table 5.2 shows the accuracy of validation, training, and testing for a classifier trained and tested on power-ENF data from seven different grid-regions, along with testing accuracy for audio record-

Table 5.1: Number of training and testing examples in the dataset.

Grid-region	Training (power)	Testing (power)	Testing (audio)
Spanish island	113	28	148
Turkey	220	55	210
Ireland	161	40	0
China	135	33	0
US east	244	61	189
US west	139	34	115
Quebec	48	12	0
Total	1060	263	662

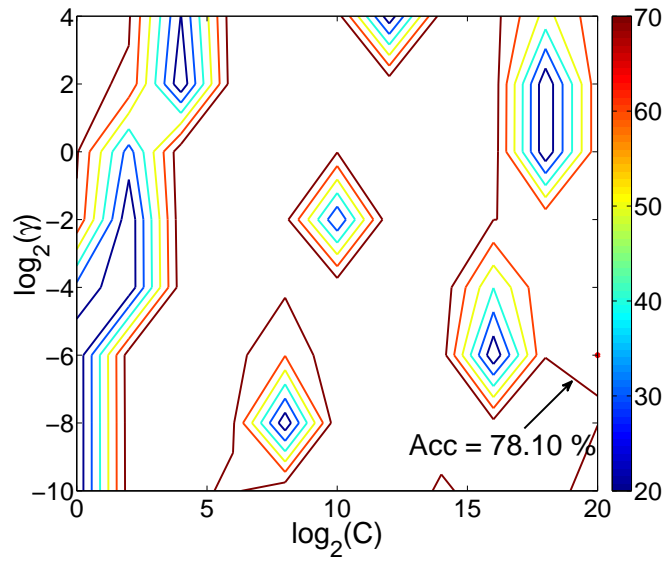


Figure 5.4: Classification accuracy of optimal parameter selection search of RBF kernel parameters  $C$  and  $\gamma$  using 10-fold cross validation.

ings from four power grids. This table also lists the testing accuracy for the audio recordings from four different grid-regions for the same classifier. From the table we observe that the testing accuracy for the power recordings compare favorably to the validation and training accuracy. These numbers indicate that the proposed location classifier can provide a good classification for ENF signals recorded from different power grids. The current study uses only the AR modeling parameters and average frequency as features for classification to investigate the effectiveness of the AR modeling in learning the grid characteristics. It may be possible to obtain a better representation of the feature vector by using other properties, such as high order moments, the range, and the sub-band features of ENF signals [39].

Fig. 5.5 shows the classification results for power testing data from different grids. From this plot, we observe that for each grid’s testing data, a maximum number of testing examples are correctly classified into that particular grid. For the North American grids, the classification accuracy is better than for the more diverse data from Europe and Asia. For example, recordings from the Quebec grid show a 100% classification accuracy. The classification accuracy for the China recordings is higher when compared with other European and Asian grids. It is worth noting that the geographical regions covered by North American and China grids are expansive compared with the Turkey, Spain, and Ireland grids. These results indicate that the proposed feature extraction method can be used to determine the grid-region of a recording, especially for power recordings.

The classification performance degrades when audio recordings are used for testing, as seen in Table. 5.2. The accuracy for the classifier trained on power-ENF



Table 5.2: Classification accuracy for power-ENF training. All numbers are shown in % terms.

Data	Validation accuracy	Training accuracy	Testing accuracy
Power	78.10	81.18	71.14
Audio(Testing only)	-	-	44.26

signals and tested on audio recordings is 44.26%, which is considerably lower than the testing accuracy of 71.14% obtained for the power testing signals. This happens because of the presence of noise in the audio recordings, which affects the ENF signal estimation. The presence of noise makes the feature extraction scheme and the classification scheme prone to error.

To understand the effect of the noise on the classification accuracy, we evaluate the performance of the proposed location classification method when the classifier is trained on noisy ENF data at different noise levels. We create the noisy dataset by adding noise to the power-ENF signal at various SNRs. We train the SVM on the feature vectors obtained from this noisy data. We use the same power-ENF dataset used in our previous experiments for the training and the testing. We test the performance of this classifier with test data at different SNRs. The plots of the accuracy for classifiers obtained using such a methodology for different training and testing SNRs are shown in Fig. 5.6. In this figure, the x-axis represents the test SNR conditions, and the lines correspond to different training conditions. From this figure, we observe that a mismatch between the training and the testing conditions leads to a lower classification accuracy compared with the cases using the same

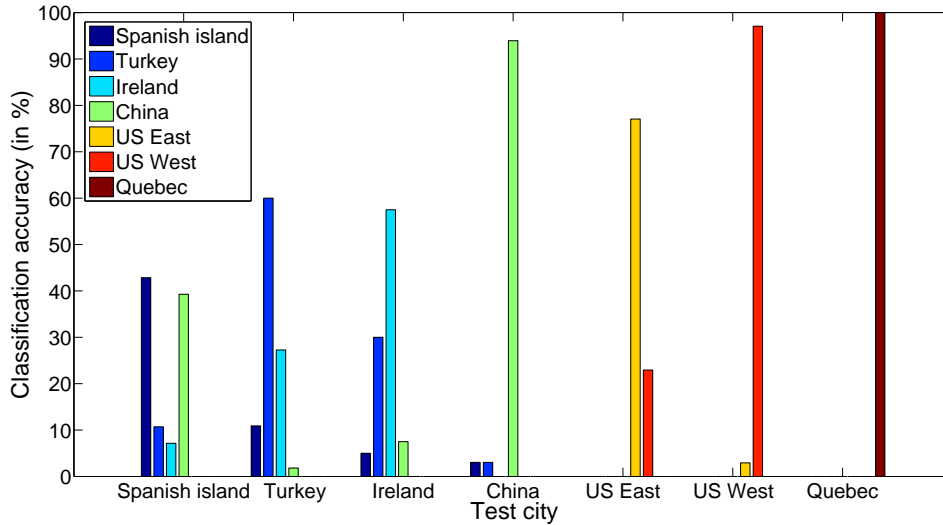


Figure 5.5: Classification accuracy on power testing data for SVM trained on power data.

training and the testing SNRs. For example, the highest testing accuracy for the classifier trained on 10dB SNR condition is 74.52% for a testing SNR of 10dB, while the accuracy drops to 25% for a testing SNR of 5dB. This suggests that the proposed location classification provides the best accuracy when the training and the testing data have matching noise conditions.

For the location classification in multimedia recordings, where the ENF signal is embedded into an audio or a video recording with the signal’s content, it is not possible to know the training conditions beforehand. For example, the SNR of ENF signals from our audio test data varies between 5dB-15dB, depending on the location-of-recording. These variations in the SNR may arise due to the nature of the recording device, sources, or mechanism of ENF embedding onto the sensor recording at the location of recording, and it may not be possible to estimate the accurate

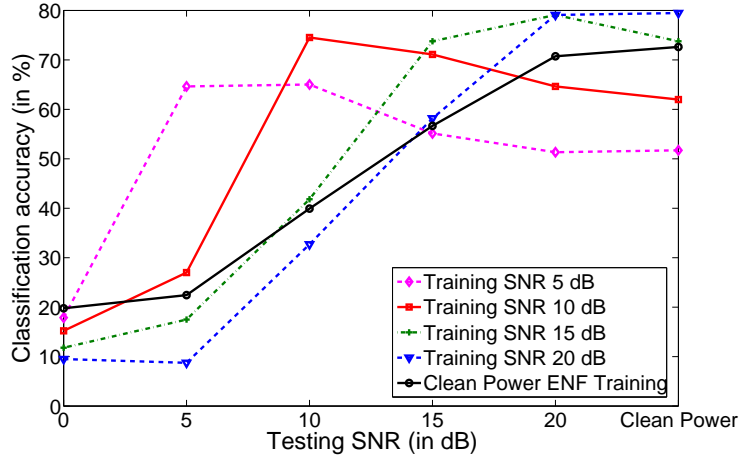


Figure 5.6: Classification accuracy for different training and testing conditions.

value of SNR for the given query multimedia recording. It, therefore, becomes necessary to develop a mechanism that can automatically adapt to different types of noise conditions. We examine a multi-conditional learning approach to adapt our classifier to different noise levels in the next section.

## 5.4 Noise Adaptation using Multi-Conditional Learning

In the previous section, we have observed that a mismatch in the training and the testing conditions can lead to a lower value of the grid of recording classification accuracy, unlike when noise conditions between the training and testing match. For example, the accuracy of a classifier trained on power data and tested on power data was 71.14%, but it reduces to 21% with test data at 10dB SNR. This may occur because of the limitations that the simple feature extraction mechanism is not robust against noise, and the classifier may be unable to adapt to the testing conditions. As

this work focusses on demonstrating the effectiveness of the AR modeling parameters in learning the grid specific characteristics, we do not delve into extracting other types of feature for better signal representation. Instead, using only the existing AR modeling parameters as features, we investigate a multi-conditional learning approach to adapt to the mismatch between the training and the testing conditions. Such multi-conditional learning has been used in the speech processing literature for the similar problem of mismatch in the noise conditions, for tasks such as speaker recognition and speech understanding [40] [41].

#### 5.4.1 System Model

Our proposed multi-conditional learning approach has a separate classification model for each of several different representative noise conditions to match a variety of environments. For each noise environment, the corresponding classification model can be derived by transforming the model parameters to better characterize the given environment. This transformation is performed based on learning the environmental model on the corresponding environment’s training data. This can be achieved by using ENF signals at different noise conditions for feature extraction, and learning separate classification models for each noise condition. Such ENF signals can be obtained either from the noisy audio signals or by adding synthetic white Gaussian noise (WGN) at different noise conditions to the clean power-ENF signals. In our work, we construct our training dataset by generating WGN at various SNR conditions and adding it to the power-ENF signal, then we perform the

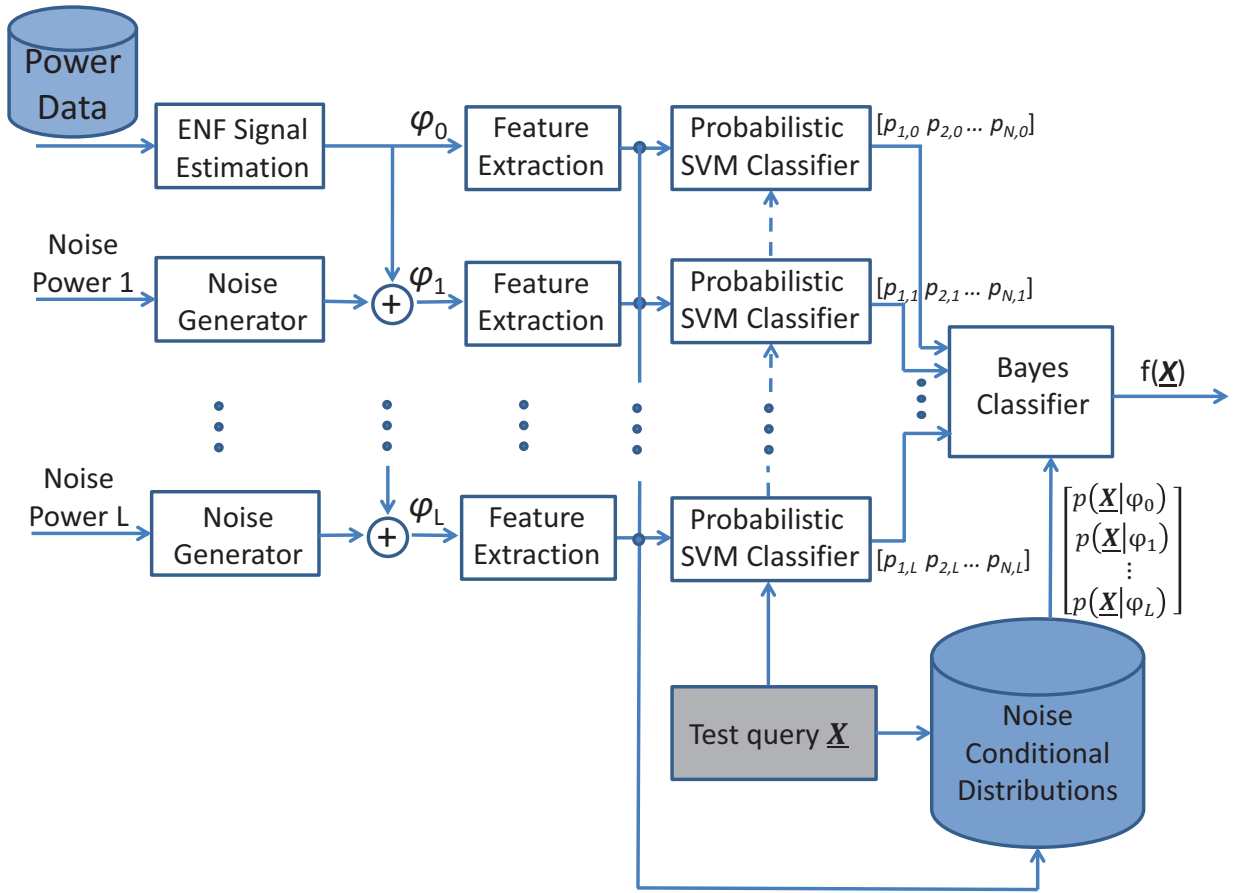


Figure 5.7: Block diagram of multi-conditional learning for noise adaptation.

feature extraction on these data.

When a test query is introduced to the classifier trained using the multi-conditional learning approach, it may find the class of the given test query in one of the two ways. Firstly, if the noise conditions can be identified before classification, the model corresponding to the identified noise conditions can be used. Alternatively and more generally, the results from all training models can be fused together to derive a classification decision criteria. We adopt a Bayesian fusion approach to compute the classification results, and this framework requires estimating the likelihood of observing the query feature vector for each noise conditions.

A block diagram of the proposed system is shown in Fig. 5.7. More specifically, let  $\phi_0$  denote the training dataset containing clean power-ENF signal, and  $p(\underline{\mathbf{X}}|\phi_0)$  represent the likelihood of observing feature vector  $\underline{\mathbf{X}}$  from  $\phi_0$ . The proposed approach for noise adaptation includes two steps. The first step is to generate multiple copies of the original training set  $\phi_0$  by introducing different noise conditions at different SNRs to the clean power-ENF signal. We arrive at an augmented training set at different noise conditions and denote the copies by  $\phi_0, \phi_1, \dots, \phi_L$ , where  $\phi_l$  denotes the  $l^{\text{th}}$  training set derived from  $\phi_0$  by adding WGN at  $l^{\text{th}}$  noise power.

Under this setting, our goal is to predict the class of grid-region of recording from the given feature vector  $\underline{\mathbf{X}}$  obtained from a test recording. The class prediction is made by first estimating the probability of  $\underline{\mathbf{X}}$  belonging to each candidate class, and assigning the class with the maximum probability as the location-of-recording. Let  $C = \{1, 2, \dots, N\}$  denote the candidate classes. Then, the probability of  $\underline{\mathbf{X}}$  belonging to class  $i$  can be written as:

$$\begin{aligned}
 p(C = i|\underline{\mathbf{X}}) &= \sum_{l=0}^L p(C = i, \phi_l|\underline{\mathbf{X}}) \\
 &= \sum_{l=0}^L p(C = i|\underline{\mathbf{X}}, \phi_l)p(\phi_l|\underline{\mathbf{X}}) \\
 &= \sum_{l=0}^L p_{i,l}(\underline{\mathbf{X}})p(\phi_l|\underline{\mathbf{X}}), \tag{5.2}
 \end{aligned}$$

where  $p_{i,l}(\underline{\mathbf{X}})$  denotes the probability of  $\underline{\mathbf{X}}$  belonging to class  $i$  under  $l^{\text{th}}$  noise condition;  $p(\phi_l|\underline{\mathbf{X}})$  denotes the probability of  $\underline{\mathbf{X}}$  belonging to the noise condition  $\phi_l$ . The expression for  $p(\phi_l|\underline{\mathbf{X}})$  is obtained by the Bayes formula as follows:

$$p(\phi_l|\underline{\mathbf{X}}) = \frac{p(\underline{\mathbf{X}}|\phi_l)p(\phi_l)}{\sum_{l'=0}^L p(\underline{\mathbf{X}}|\phi_{l'})p(\phi_{l'})}, \tag{5.3}$$

where  $p(\phi_l)$  denotes the prior probability of  $l^{\text{th}}$  noise condition in the dataset. Bringing Eq. (5.3) into Eq. (5.2), we arrive at:

$$\begin{aligned} p(C = i|\underline{\mathbf{X}}) &= \sum_{l=0}^L p_{i,l}(\underline{\mathbf{X}})p(\phi_l|\underline{\mathbf{X}}) \\ &= \sum_{l=0}^L p_{i,l}(\underline{\mathbf{X}}) \frac{p(\underline{\mathbf{X}}|\phi_l)p(\phi_l)}{\sum_{l'=0}^L p(\underline{\mathbf{X}}|\phi_{l'})p(\phi_{l'})}. \end{aligned} \quad (5.4)$$

Assuming that the prior  $p(\phi_l)$  is uniform for all  $l$ , the expression in Eq. (5.4) can be written as:

$$p(C = i|\underline{\mathbf{X}}) \propto \sum_{l=0}^L p_{i,l}(\underline{\mathbf{X}})p(\underline{\mathbf{X}}|\phi_l). \quad (5.5)$$

The decision rule of the class assignment for query vector  $\underline{\mathbf{X}}$  can be written as:

$$\begin{aligned} f(\underline{\mathbf{X}}) &= \operatorname{argmax}_{i=1,2,\dots,N} p(C = i|\underline{\mathbf{X}}) \\ &= \operatorname{argmax}_{i=1,2,\dots,N} \sum_{l=0}^L p_{i,l}(\underline{\mathbf{X}})p(\underline{\mathbf{X}}|\phi_l). \end{aligned} \quad (5.6)$$

To perform the classification shown in Eq. (5.6), we compute  $p_{i,l}(\underline{\mathbf{X}})$  and  $p(\underline{\mathbf{X}}|\phi_l)$ . The value of  $p_{i,l}(\underline{\mathbf{X}})$  is obtained by using a probabilistic SVM approach and training separate classifiers under each of the  $L + 1$  signal conditions [42]. The probabilistic SVM assigns a likelihood of each feature vector  $\underline{\mathbf{X}}$  belonging to class  $i$  based on the distance of each class boundary from feature vector [43]. A discussion on probabilistic SVM is provided in Appendix at the end of this chapter. The value of  $p(\underline{\mathbf{X}}|\phi_l)$  is obtained by learning the distribution of  $\underline{\mathbf{X}}$  for each noise condition. We learn these conditional distributions of  $\underline{\mathbf{X}}$  using two widely used distribution learning techniques, namely Gaussian mixture models (GMMs) and Parzen-window density estimation.

a) *Gaussian Mixture Models*: Gaussian mixture models (GMMs) are considered universal approximations of densities, i.e., given a sufficient number of mixture components, they can approximate any distribution [44]. A GMM is the weighted sum of  $M$  component densities, each of which is Gaussian. Mathematically, it can be written as:

$$p(\underline{\mathbf{X}}|\phi_l) = \sum_{i=1}^M w_i^{(l)} \mathcal{N}(\underline{\mu}_i^{(l)}, \underline{\Sigma}_i^{(l)}), \quad (5.7)$$

where  $\mathcal{N}(\underline{\mu}_i^{(l)}, \underline{\Sigma}_i^{(l)})$  represents a Gaussian probability density function (pdf) with mean vector  $\underline{\mu}_i^{(l)}$  and covariance matrix  $\underline{\Sigma}_i^{(l)}$  for the  $l^{th}$  noise condition. Given the training feature vectors  $\underline{\mathbf{X}}$  from all training examples of the  $l^{th}$  noise condition, we aim to learn the parameters values  $\{w_i^{(l)}, \underline{\mu}_i^{(l)}, \underline{\Sigma}_i^{(l)}\}$  of all  $M$  components of the GMM.

We use the popular expectation-maximization (EM) algorithm to obtain the maximum likelihood estimates of parameters in these probabilistic models, where the models depend on unobserved latent variables [45]. EM alternates between performing two steps. In the expectation (E) step, it computes the expectation of the likelihood function by including the current estimate of parameters and the latent variables as if they were observed; In the maximization (M) step, it computes the maximum likelihood estimates of the parameters that maximize the expected likelihood function found on the E step. The parameters found in the M step are then used for another E step, and the process repeats until a criterion for convergence is reached.

b) *Parzen-window Density Estimation*: The Parzen-window density estimation is



another popular probability density estimation method based on the observed samples of a variable [46]. This method estimates the density by superimposing kernel functions at each observation of the data. In this way, each observation of the variable contributes to the density estimation.

Let  $(x_1, x_2, \dots, x_n)$  be independent and identically distributed (iid) samples drawn from some distribution with an unknown density  $p(x)$ . The Parzen-windowing method estimates  $p(x)$  by the following equations:

$$\begin{aligned} p(x) &= \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \end{aligned} \quad (5.8)$$

where  $K(\cdot)$  is the kernel function satisfying the constraint that it integrates to one, i.e.,  $\int_{\mathcal{R}} K(x) dx = 1$ ;  $h$  is a window width parameter corresponding to the width of the kernel. The value of  $h$  is typically chosen as inversely proportional to the number of samples. We use a Gaussian kernel for the Parzen-window density estimation, and estimate the density for each feature separately. Assuming that each feature is an independent random variable, we obtain the density of the query feature  $\underline{\mathbf{X}}$  by multiplying the densities of each feature as following:

$$p(\underline{\mathbf{X}}|\phi_l) = p(f_{ave}|\phi_l)p(a_1|\phi_l)p(a_2|\phi_l)p(\sigma_v^2|\phi_l). \quad (5.9)$$

Fig. 5.8 shows the densities for the normalized innovation energy feature,  $\sigma_v^2$ , under different noise conditions and estimated using the Parzen-windowing method. The normalization results in the mean and the variance of  $\sigma_v^2$  as 0 and 1, respectively. From this figure, we observe that the densities for different SNRs have different

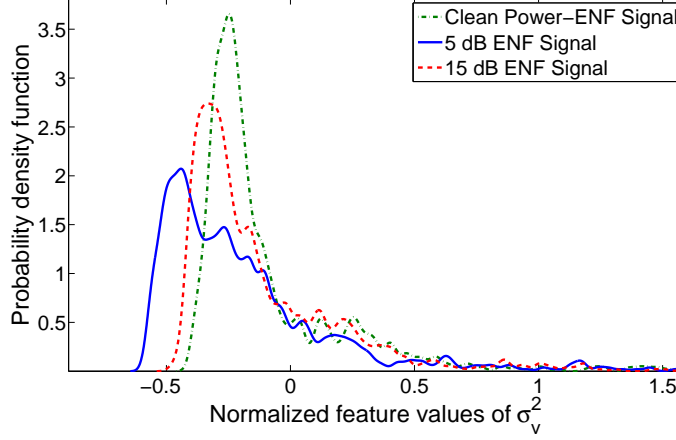


Figure 5.8: Parzen-window density estimation for different noise conditions for normalized innovation energy feature.

distributions, and our Bayesian fusion based classification method takes these noise condition distributions into account.

*Max Probability Approach:* We also attempt an alternative approach to the Bayesian approach to evaluate the location classification accuracy of the proposed method. In this method, we train  $L + 1$  probabilistic classifiers under each noise condition, similar to the Bayesian fusion based approach. In order to reach the final decision criteria, the output of each classifier is used to select for a candidate location class based on its output probabilities. More specifically, for  $l^{th}$  classifier, the vote count for  $i^{th}$  candidate location increases by one, if  $p_{i,l}(\underline{\mathbf{X}}) > p_{j,l}(\underline{\mathbf{X}}), \forall j \neq i$ . After the votes of each classifier are registered, the class obtaining the highest quantity of votes is declared as the grid-region of recording. This approach may benefit from the multi-conditional training, and it does not require any explicit conditional distribution for noise.

## 5.4.2 Results and Discussions

We use the proposed noise adaptation framework, described in Section 5.4.1 and shown in Fig. 5.7, to evaluate the location classification accuracy on the same dataset used to evaluate the classifier described in Section 5.3.2. We train multiple probabilistic SVM based classifiers for clean power-ENF signals and at five different noise levels of 0dB, 5dB, 10dB, 15dB, 20dB, respectively. The noisy signals are obtained after adding noise at the corresponding energy to the power-ENF signals. We also learn the distribution  $p(\underline{\mathbf{X}}|\phi_l)$ , for the feature vector  $\underline{\mathbf{X}}$  under each of the noise conditions.

We test the classification accuracy of the proposed classifier for a variety of different testing conditions. The plots of the classification accuracy for the clean power-ENF and the audio-ENF testing data, when GMMs are used to estimate the noise conditional distributions are shown in Fig. 5.9. From this figure, it can be observed that the classification accuracy increases as the number of mixture models used to train the GMM for noise conditional distribution is increased. The optimal number of mixtures is determined to be  $M = 12$ , where the classification accuracy for the clean power-ENF signal and the audio-ENF signal is the highest. Fig. 5.10 shows the classification accuracy under different test SNRs when the noise conditional distributions of  $\underline{\mathbf{X}}$  are learned using different number of mixtures. From this figure, it can be observed that the classification accuracy improves as the test SNR increases. This contrasts to Fig. 5.6, in which the classification accuracy was highest when the testing and the training conditions were matching. We again

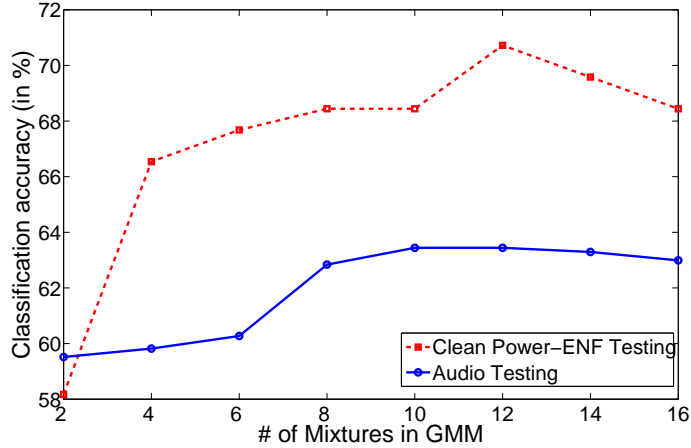


Figure 5.9: Classification accuracy under the proposed Bayesian framework when the noise conditional distributions are learnt using Gaussian mixture models.

observe that  $M = 12$  mixtures provides the highest classification accuracy for the test conditions of 5dB, 15dB, 20dB, and clean power-ENF signal. We use this optimal value of mixture numbers in our subsequent results.

In Fig. 5.11, we compare the classification accuracy of the GMM based Bayesian method with the Parzen-windowing-based method. From this figure, we observe that the GMM based method performs slightly better than the Parzen-windowing method, at test SNR levels of higher than 15dB. For low SNR conditions, the improvement in classification accuracy when using the GMM based method is more significant than that of the Parzen-windowing method. We also observe that the maximum probability approach shows a bias towards 15dB test SNR, and the classification accuracy for this test condition is higher than that of the two Bayesian fusion approaches. The classification accuracy for all location classification methods described in this chapter are summarized in Table 5.3 for power and audio signals,

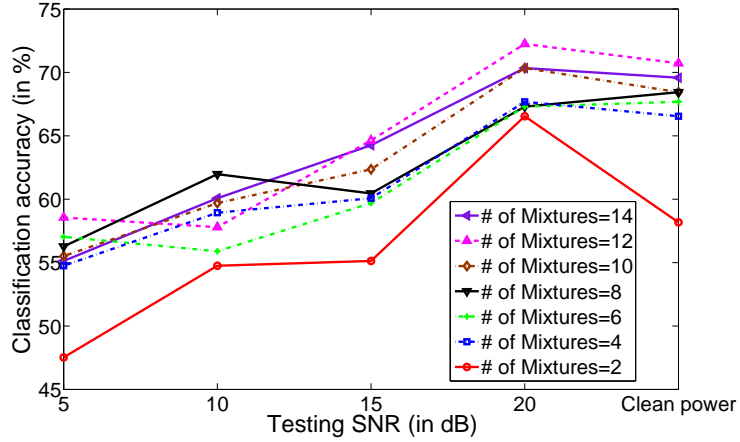


Figure 5.10: Classification accuracy for different test conditions under the proposed Bayesian framework when the noise conditional distributions are learned using Gaussian mixture models.

whereby the classifier is trained on the power-ENF data. From this table, we see that the highest value of the classification accuracy on audio data occurs with the GMM based Bayesian classifier. The improvement in the audio classification accuracy improves from 44.26%, when no noise adaptation is performed, to 63.4% for GMM based Bayesian classification. From these results, we conclude that the proposed Bayesian approach to noise adaptation is promising to improve the location classification accuracy of audio recordings.

The proposed Bayesian classification approach provides a measure of the probability of feature vector  $\underline{\mathbf{X}}$  belonging to  $i^{th}$  class in Eq. (5.4), which estimates the grid-region of recording using the decision rule in Eq. (5.6). The decision rule locates the class with the maximum probability and assigns the corresponding class as the grid of recording.

Table 5.3: Classification accuracy on power-ENF training for all the methods. All numbers are in shown in % terms.

Testing data	No noise adaptation	Max prob approach	Parzen windowing	GMM $M = 12$
Power	71.14	61.98	68.82	70.72
Audio	44.26	56.49	62.23	63.4

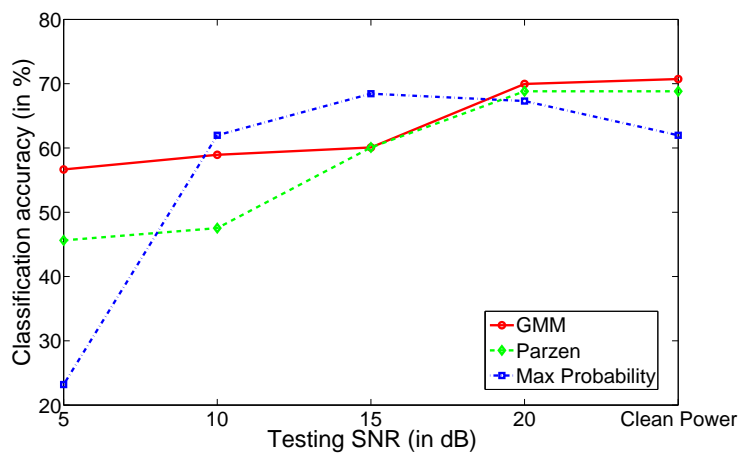
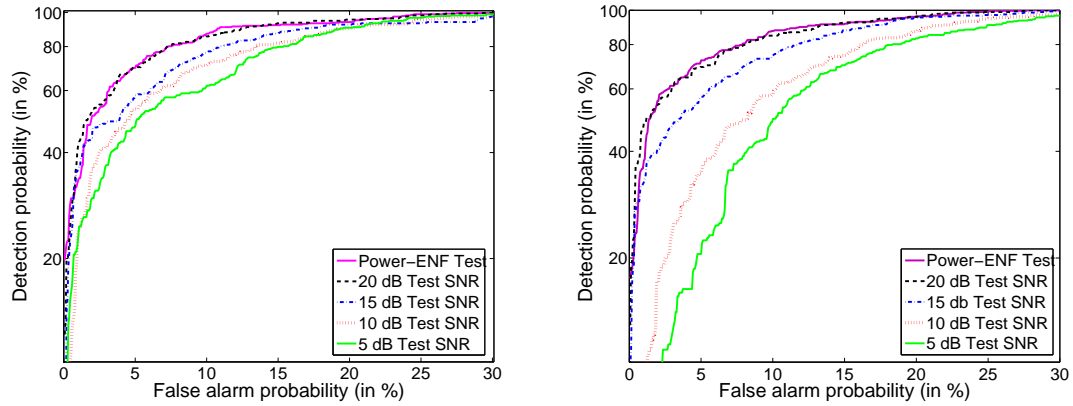


Figure 5.11: Comparison of classification accuracy for GMM, Parzen-windowing, and max probability approach.

We also examine a related problem of the location verification, where we determine a binary answer to verify if the given recording was conducted in the claimed grid. The information about the claimed grid can be embedded with a recording in GPS enabled devices. For this purpose, we use the proposed Bayesian fusion approach and obtain the probability of the query feature vector  $\underline{\mathbf{X}}$  belonging to the claimed location in the metadata. If the  $i^{th}$  location is the claimed recording grid, the probability  $p(C = i|\underline{\mathbf{X}})$  can be obtained using Eq. (5.4). We use a threshold detector to compare the value of  $p(C = i|\underline{\mathbf{X}})$  with a threshold  $\tau$ , and decide whether the claimed location as correct if this value is greater than  $\tau$ , i.e,

$$\delta_L(p(C = i|\underline{\mathbf{X}})) = \begin{cases} \text{Claimed location is correct, if } p(C = i|\underline{\mathbf{X}}) > \tau, \\ \text{Claimed location is wrong, if } p(C = i|\underline{\mathbf{X}}) \leq \tau \end{cases} \quad (5.10)$$

We measure the localization performance of the proposed detector in terms of its receiver operating characteristics (ROC). We use the same dataset used in the previous experiments. We assume that each recording test example has seven copies with each containing one of the seven candidate locations in the metadata. The plots of the ROC for the GMM and the Parzen-windowing based method under this setting for different testing conditions are shown in Fig. 5.12(a) and 5.12(b), respectively. From these figures, we observe that the ROC performance improves as the test SNR is increased. The location verification detector provides a better measure of the confidence as compared with the previous experiments on grid-region classification. For example, for clean power-ENF test condition, a threshold  $\tau$  can be determined, which provides 87% chances of detection for a false alarm rate of 10%,



(a) GMM based Bayesian method,  $M = 12$       (b) Parzen-window based Bayesian method

Figure 5.12: Receiver operating characteristics of the proposed location verification detector.

for both the methods. We also observe that the ROC performance of the GMM based method is better than the Parzen-windowing method for the test SNRs of 5dB and 10dB, similar to the classification accuracy observed in Fig. 5.11.

## 5.5 Chapter Summary

In this chapter, we have explored a novel application of the ENF signal analysis in grid-region of recording estimation of multimedia recordings. We have demonstrated the characteristics of the ENF signals in different power grids. We have used AR model proposed in Chapter 4 to capture these variations. Given these observations, we have proposed a location classifier based on AR parameters and average frequency as the features to learn a SVM based classifier on the power-ENF data from seven grids. We have also demonstrated the effect of mismatch in training



and test noisy conditions on the classification accuracy. To compensate the effect of mismatch, we propose a Bayesian fusion approach, combining SVM classifiers trained under different noisy conditions. The proposed approach can automatically adapt to various noise conditions. Using this approach, we can obtain a location classification accuracy of approximately 64% for audio data, as compared with the location classification accuracy of approximately 44% with no noise adaptation. In the next chapter, we explore the problem of identifying the area of recording within the same grid using another set of ENF signal characteristics.

## Appendix: Probabilistic Support Vector Machines

We employ the probabilistic SVM framework proposed in [42] to compute the probability  $p_{i,l}(\underline{\mathbf{X}})$  that a given test vector  $\underline{\mathbf{X}}$  comes from  $i^{th}$  class under  $l^{th}$  training noise condition. Without loss of generality, we provide a discussion on estimating  $p_i(\underline{\mathbf{X}})$  by dropping the subscript  $l$  and argument  $\underline{\mathbf{X}}$ . With the assumption that the class conditional probabilities  $p(\underline{\mathbf{X}}|C = i)$  is exponentially distributed [43], the estimates  $r_{ij}$  of the pairwise class probabilities  $r_{ij} = p(C = i|C = i \text{ or } j, \underline{\mathbf{X}})$  can be found by the following parametric fitting:

$$r_{ij} = \frac{1}{1 + e^{A\hat{g}+B}},$$

where  $A$  and  $B$  are estimated by minimizing the negative log-likelihood function, and  $\hat{g}$  are the decision values on training data. We then find  $p_i = p(C = i|\underline{\mathbf{X}})$ , the probability that the data sample comes from the  $i^{th}$  class for a  $N$ -class SVM, by

solving the following optimization problem:

$$\begin{aligned} \min_{p_1, p_2, \dots, p_N} \quad & \sum_{i=1}^N \left( \sum_{j, j \neq i} r_{ji} p_i - r_{ij} p_j \right)^2 \\ \text{subject to} \quad & \sum_{i=1}^N p_i = 1, \quad p_i \geq 0, i = 1, 2, \dots, N. \end{aligned}$$

Further details of the method can be found in [42], and the implementation of probabilistic SVM used in this thesis can be found at [38].

# Chapter 6

## Intra-grid Location Estimation using ENF Signals

### 6.1 Chapter Introduction

The potential of Electrical Network Frequency signals in multimedia forensics has been mainly explored for timestamping and tampering detection. In chapter 5, we demonstrated that the statistical modeling parameters of ENF signals can be used to locate the grid-region of the recording. An important, intriguing question continues to be: “Can the ENF signal be used to estimate or verify the place of recording of an audio or a video recording within the same grid-region?” An affirmative answer could potentially be used in many applications such as national defense and security, and automatic region-tagging of multimedia data on social networking sites like YouTube, Facebook, and Twitter.

As shown in Chapter 5, it is possible to differentiate between the recordings

in different grids at an inter-grid level, as the fluctuations in ENF signals typically vary at the same time across independently operated grids. At an intra-grid level, most existing works have assumed that ENF signals across different locations in an interconnected power grid are similar at the same time. However, minor variations are likely to be present in the frequency fluctuations at different locations given local changes in the load and the finite propagation speed of the effects of such load changes have on other parts of the grid [47]. In this chapter, we study such effects by conducting experiments on the ENF data collected from several locations within the US east grid. It will be shown later in the chapter that differences exist among simultaneous ENF signals extracted from recordings taken in various locations within the same interconnected power grid. Our study builds a foundation to design a localization protocol, based on a method of half-plane intersection, to estimate the location of recordings within the same grid-region. In this study, we also discuss the challenges to localization arising from the noisy nature of the ENF signal from multimedia recordings.

## 6.2 Propagation Mechanism of ENF Signal

The fluctuations in ENF signals in the same grid happen because of the dynamic nature of the grid load. Power demand and supply in a given area follows a cyclic pattern. For example, demand increases during evening hours in a residential neighborhood, as people switch on air-conditioning and other power units. For robust operation, any load change within a grid is regulated [6]. An increase in the

load causes the supply frequency to drop temporarily; the control mechanism senses the frequency drop and starts drawing power from adjoining areas to compensate for the increased demand. As a result, the load in adjoining areas also increases, which leads to a drop in the instantaneous supply frequency. The overall power supply will elevate to compensate for the rising load, which leads to a drop in the instantaneous supply frequency in those regions. A similar mechanism compensates for an excess supply of power flow that leads to surges in the supply frequency.

A small change in the load in a given area may have a localized effect on the ENF in that area. However, a large change such as the one caused by a generator failure may have affect an entire grid. In the US east grid, these changes are known to propagate along the grid at a typical speed of approximately 500 miles per seconds [47]. We conjecture that load change may introduce location specific signatures in the ENF patterns, and such differences may be exploited to narrow the location of a recording within the grid. Due to the finite speed of propagation of frequency disturbances in the grid, we anticipate that ENF signals could have greater similarity for locations close to each other as opposed to those further apart. Such a property of ENF signal propagation across the grid can potentially be used for localization at a finer resolution within a grid by comparing the similarity of the ENF signal in question with ENF databases that may be available for a set of anchor locations within that grid.

## 6.3 Location Dependence of ENF Signals

As a first step to explore the availability of location dependent properties of ENF signals, we focus on the ENF signal obtained directly from the power mains. This provides the most favorable conditions in terms of a high signal-to-noise ratio (SNR) of the power-ENF signal. ENF signals collected across different locations are similar to each other over time, so exploration using high SNR signals may enhance understanding if ENF signals exhibit location specific characteristics, which can potentially be exploited to devise a localization protocol. Such a study may be considered an initial step toward gaining an understanding of the location estimation capabilities of ENF signals, leading to solutions to the more difficult problem of location estimation from audio and video recordings; ENF signals in such recordings are present in a distorted form and at a very low SNR.

Fig. 6.1(a) shows a plot of ENF signals extracted from three simultaneous short recordings conducted in College Park-MD, Princeton-NJ, and Atlanta-GA. These three areas are located in the US east grid. From this figure, we observe that all three ENF signals correlate strongly at a macroscopic level; however, in the enlarged plot shown in Fig. 6.1(b), differences become noticeable across the three recordings. We extract these variations using a filtering mechanism, and then compare them to gain an understanding of the relationship between signals recorded at different locations.

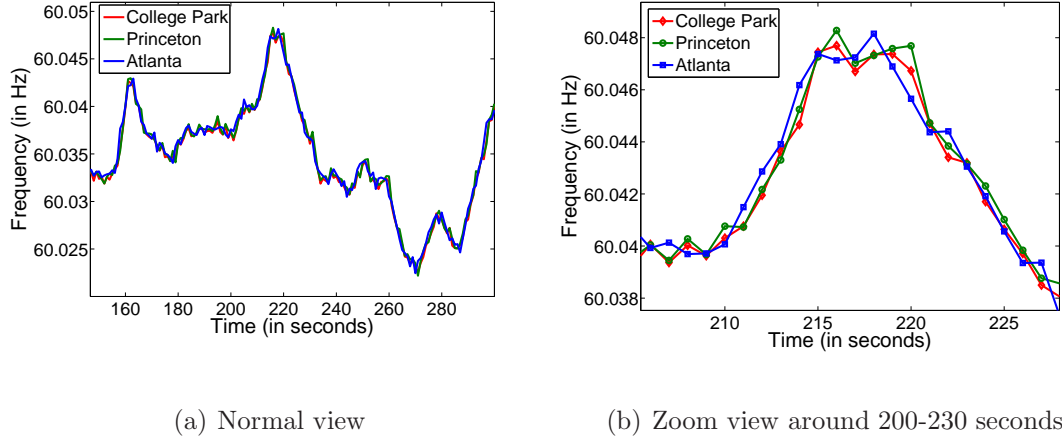


Figure 6.1: Sample ENF signals extracted from recordings done in three locations in the US eastern grid at the same time. (Figures are best viewed in colors).

### 6.3.1 Signal Processing Mechanism to Extract ENF

#### Variations

As seen in Fig. 6.1(b), variations occur at high frequencies in simultaneous ENF signals recorded across different locations of the same grid. To extract these variations, we use a high pass filtering mechanism by passing temporally aligned ENF signal,  $f^{\{k\}}(n)$ , recorded at  $k^{th}$  location through a smoothing filter, and subtract the resulting output signal from  $f^{\{k\}}(n)$ . The corresponding high pass filtered output,  $f_{hp}^{\{k\}}(n)$  is given by:

$$f_{hp}^{\{k\}}(n) = f^{\{k\}}(n) - \sum_{m=-\frac{M-1}{2}}^{\frac{M-1}{2}} w(m)f^{\{k\}}(n-m), \quad (6.1)$$

where  $f^{\{k\}}(n)$  is the ENF value at time  $n$ ,  $w(\cdot)$  is the coefficient of the smoothing filter, and  $M$  is the filter order for feature extraction, chosen as an odd number. After extracting high pass filtered signals for each location, their pair-wise cross-

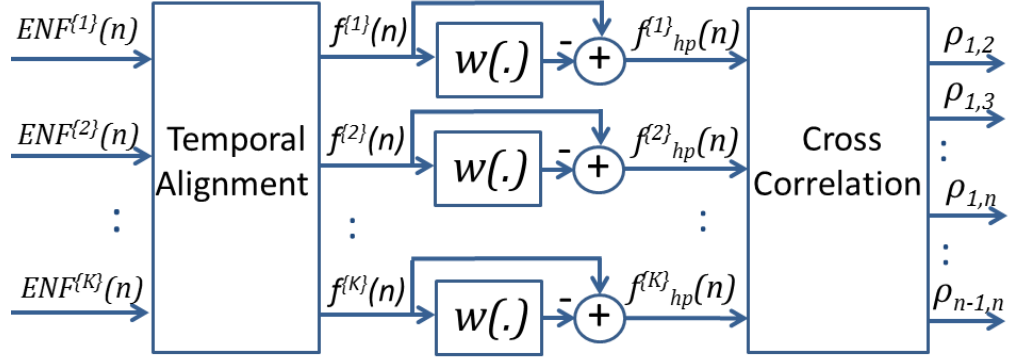


Figure 6.2: Signal processing mechanism to extract intra-grid ENF Signatures.

correlations are obtained. The pair-wise cross-correlation between any two filtered segments at time  $n$  from the  $k^{th}$  and the  $l^{th}$  location is given by:

$$\rho_{k,l} = \frac{\sum_{p=0}^{N-1} f_{hp}^{\{k\}}(n+p) f_{hp}^{\{l\}}(n+p)}{\sqrt{\sum_{p=0}^{N-1} (f_{hp}^{\{k\}}(n+p))^2} \sqrt{\sum_{p=0}^{N-1} (f_{hp}^{\{l\}}(n+p))^2}}, \quad (6.2)$$

where  $N$  is the length of the signal segment. A block diagram representing this signal processing mechanism is shown in Fig. 6.2.

### 6.3.2 Case Study 1: 3-Location Data on the US East Coast

In this section, we describe our experiments on a 10-hour long simultaneous recording of power data from three locations in the US east grid: College Park in Maryland, Princeton in New Jersey, and Atlanta in Georgia. We use the mechanism described in Section 6.3.1 to estimate the cross-correlation between filtered ENF data from all three locations. We divide the signal into non-overlapping segments of 10 minutes each. Instantaneous frequency is estimated every 1 second using the subspace based ESPRIT [48] method, which provides better frequency estimation accuracy than other methods [11]. The plot of the correlation coefficients between



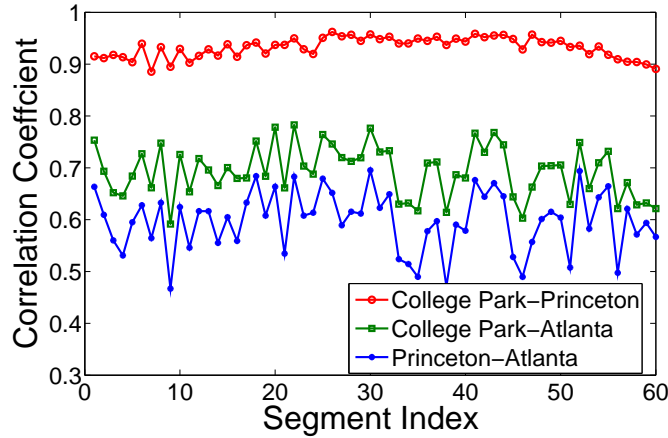


Figure 6.3: Correlation coefficient between processed ENF signals for 3-location data in US east coast for a 10-minute long query segment.

processed ENF signals at different locations for filter order  $M = 3$  is shown in Fig. 6.3. This figure shows that the correlation coefficient between the signals from city pairs far apart in geographical distance is less than that between the signals from the closer city pairs. The correlation coefficient is approximately proportional to the distance between the cities. These three cities lie in approximately a straight line on a map, as shown in Fig. 6.4. Based on these observations, we derive a relationship between the correlation coefficient of the data from different locations and their geographical distances.

Let us denote Princeton-NJ by city 1, College Park-MD by city 2, and Atlanta-GA by city 3. Assuming that city distance follows a linear relationship with the correlation coefficient, we use the values of correlation coefficients  $\rho_{1,2}$ ,  $\rho_{2,3}$ , and corresponding city geographical distances  $d_{1,2}$ ,  $d_{2,3}$  to obtain an estimate of  $d_{1,3}$  for a given observation of  $\rho_{1,3}$ . Based on the linear relationship, an estimate of  $\hat{d}_{1,3}$  for

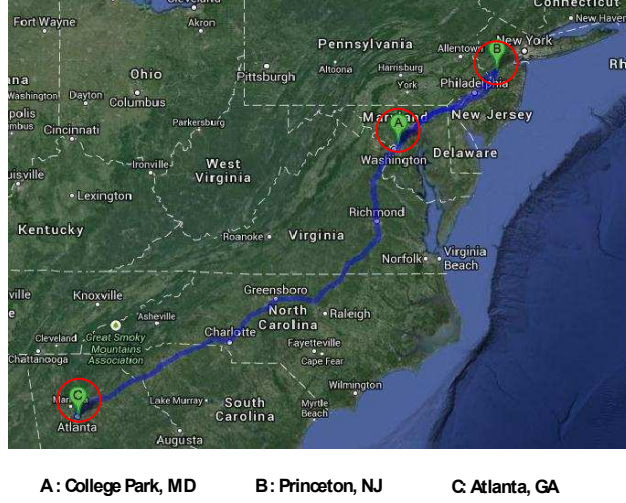
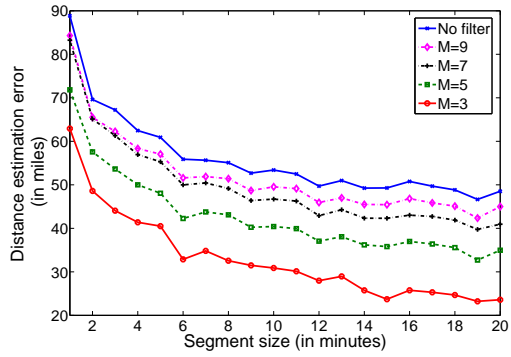


Figure 6.4: Three locations shown on a map for Case Study 1.

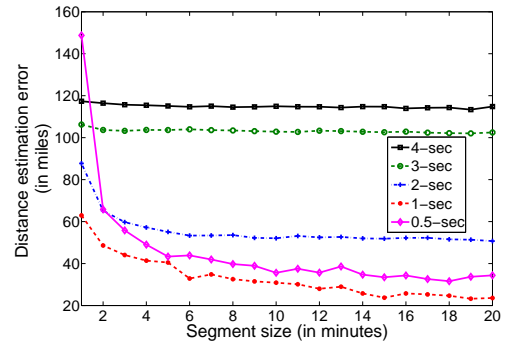
a given  $\rho_{1,3}(n)$  can be given as:

$$\hat{d}_{1,3} = d_{1,2} + \frac{d_{1,2} - d_{2,3}}{\rho_{1,2} - \rho_{2,3}}(\rho_{1,3} - \rho_{1,2}). \quad (6.3)$$

We compute the mean distance estimation error by averaging the absolute difference between  $\hat{d}_{1,3}(n)$  and true geographical distance  $d_{1,3}$  for different query segments. The plot of the mean distance error in distance estimation for different segment lengths and filter orders is shown in Fig. 6.5(a). According to this figure, when the filter order  $M = 3$  is used, a 15-minutes long segment can provide an estimate of distance within an accuracy of about 24 miles. Increasing the filter order degrades the distance estimates because the use of more data in filtering of the ENF signal averages the effects propagated due to the finite propagation speed of the frequency disturbances across the grid. To understand the effect of temporal resolution in distance estimation, we fix  $M = 3$  and plot the average distance estimation error for different durations of instantaneous frequency estimation in



(a) For different  $M$



(b) Different segment duration,  $M = 3$

Figure 6.5: Mean error in distance estimation between Princeton and Atlanta using a linear relationship between correlation coefficients and distance between the cities.

Fig. 6.5(b). From this figure, we observe that the best estimates are obtained when instantaneous frequency is estimated every 1 second. Such a phenomenon can be explained by the finite speed of signal propagation, which is empirically determined to be in the order of  $\approx 500$  miles for the US east grid [47]. As we increase the duration of data for instantaneous frequency estimation, the effect of the signal propagation averages, leading to a decrease in the accuracy of distance estimates. Decreasing the signal duration for instantaneous frequency estimation by less than 1 second leads to an error in frequency estimation itself due to the small number of data samples available for frequency estimation.

Based on this case study, we see that ENF signals have the potential to be used as location-stamps. The correlation coefficient between the data recorded at an unknown location and a known location can be used to estimate the distance of the recording location from the known location. Known locations can behave as

anchor nodes in designing localization protocols [49]. In the next section, we discuss another case study on 5-location data in the US east grid that reveals additional challenges.

### 6.3.3 Case Study 2: 5-Location Data on the US East Coast

For this experiment, power data was collected from two additional locations, Champaign in Illinois and Raleigh in North Carolina. The location of the five locations are designated on a map in Fig. 6.6. This 5-location data is four hours in duration. After temporally aligning the signals, we use the mechanism described in Section 6.3.1 to estimate the correlation coefficients between the data from different city pairs for a filter order  $M = 3$  and a segment length of 10 minutes. The plots of the correlation coefficients between data from different cities are shown in Fig. 6.7. From these figures, we observe that the correlation coefficients between the data collected from cities closer to each other are higher than from cities further apart, similar to the 3-location data. The relative magnitude of the correlation coefficients are roughly inversely proportional with the geographical distance between the cities. For example, the distance between College Park and Princeton is the least mutual distance among all city pairs, and the correlation coefficient between the data collected is the highest. However, due to a different grid density and 2-dimension relation of the electricity flows, it may not be possible to use the straight line assumption as done in Section 6.3.2.

Fig. 6.8 shows a grid density map of the US east interconnection grid, from

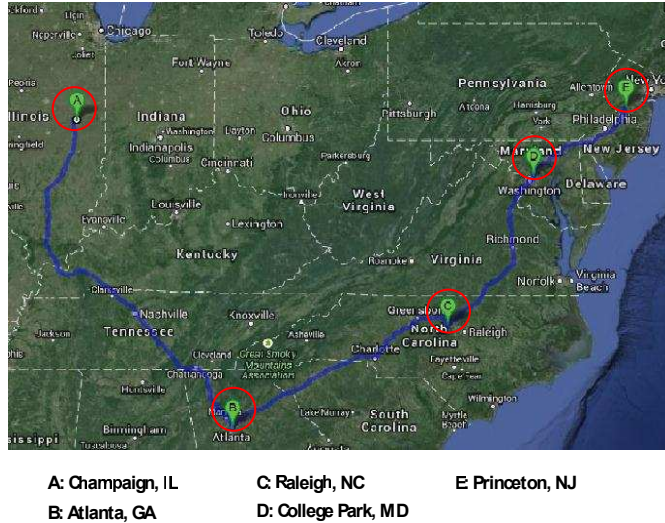
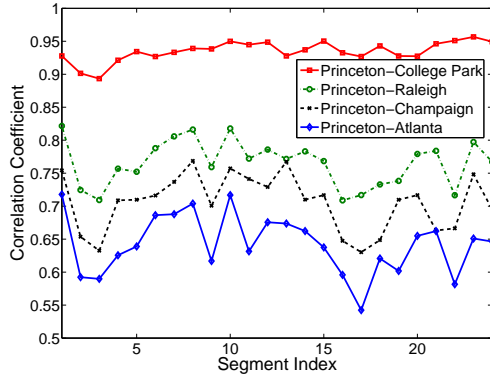
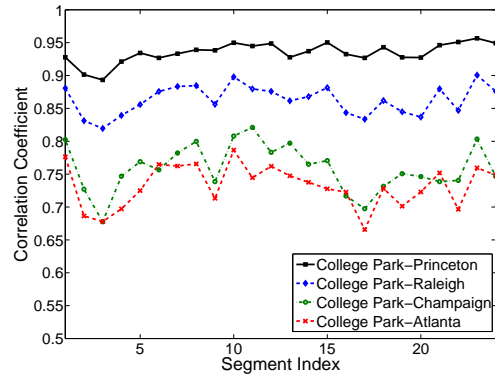


Figure 6.6: Five locations shown on a map for Case Study 2.

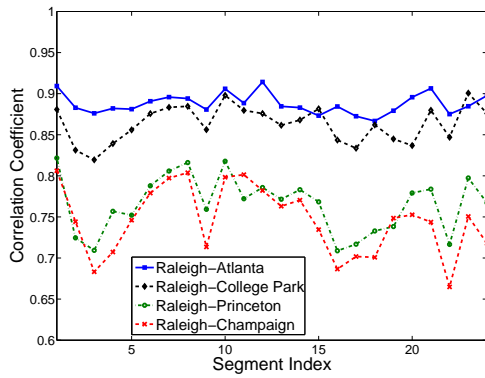
which we observe that the grid-density is non-uniform at different places and along different directions. As the flow of ENF signals over the wire lines depends on such parameters as grid topology (road distance may not be the same as the actual wire distance), grid density, etc., the correlation coefficient between data from different locations may have a complex relationship with the distance between these locations. Limited amount of data is available to us, so we must design a localization protocol without learning an explicit relationship between the correlation coefficient and the distance between different locations. Instead, we use the observation from our experiments that the pair-wise correlation between the locations far apart is less than of the pair-wise correlation between the closer cities. Using such observations, we devise a method of half-plane intersection to estimate an unknown location of recording.



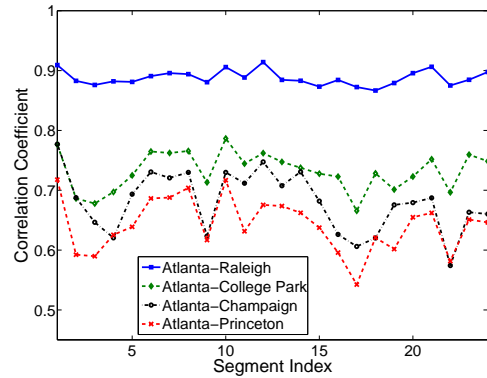
(a) Correlation with Princeton data



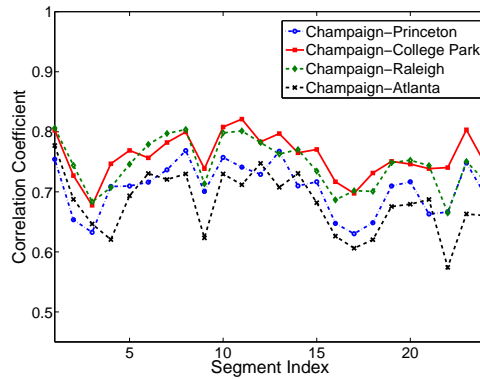
(b) Correlation with College Park data



(c) Correlation with Raleigh data



(d) Correlation with Atlanta data



(e) Correlation with Champaign data

Figure 6.7: Correlation coefficient between the processed ENF signals across different locations for 10-minute long query segment for 5-location data in the US east grid.

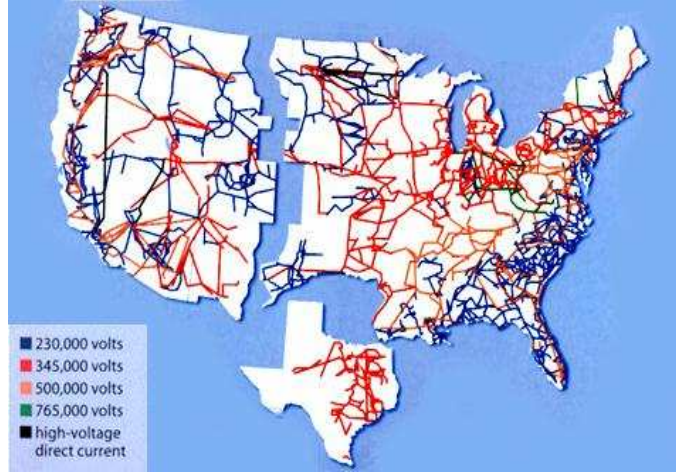


Figure 6.8: Grid density of the US east interconnection grid. [Source: <http://www.slipperybrick.com/2009/03/worm-virus-could-bring-down-us-power-grid/>]

## 6.4 Half-Plane Intersection for Localization

Let us denote the location of  $K$  anchor cities by  $P_1 = \{x_1, y_1\}, P_2 = \{x_2, y_2\}, \dots, P_K = \{x_K, y_K\}$ . Suppose we are given ENF data collected at all anchor cities, along with their known locations. Based on this information, we derive a localization protocol to estimate the unknown location of a city (denoted by  $P_{Query}$ ), based on the ENF data recorded at that location. We assume that the query city location  $P_{Query}$  and the locations of all anchor nodes lie in a rectangular region surrounding  $P_1, P_2, \dots, P_K$  and denoted by  $D$ . We refer to  $D$  as the domain of the localization region. As discussed in Section 6.3.3, if the distance between  $P_i$  and  $P_{Query}$  is greater than the distance between  $P_j$  from  $P_{Query}$ , we generally have  $\rho_{j,Query} > \rho_{i,Query}$ . Based on this observation, we claim that the estimated  $\hat{P}_{Query}$  lies in the half-plane

given by the region denoted by the set of locations  $\widehat{P}_{i,j}$  for which the following equations hold:

$$\widehat{P}_{i,j} = \begin{cases} X : \|X - P_i\|_2 > \|X - P_j\|_2, X \in D, & \text{if } \rho_{j,Query} - \rho_{i,Query} > 0, \\ X : \|X - P_i\|_2 \leq \|X - P_j\|_2, X \in D, & \text{if } \rho_{j,Query} - \rho_{i,Query} \leq 0. \end{cases} \quad (6.4)$$

The conditions described in Eq. (6.4) are the sign bit of the difference between the correlation coefficients, and these use highly quantized information from the correlation coefficient. The conditions also provide us with hard decision boundaries for the half-plane and do not take into account the noisy nature of pair-wise correlation coefficients. For example, when the correlation coefficients of the query city's ENF signal with  $i^{th}$  and  $j^{th}$  locations are close to each other, i.e., if  $|\rho_{i,Query} - \rho_{j,Query}| < \epsilon$  for a small  $\epsilon$ , the confidence in assigning a half-plane to the feasible solution set  $\widehat{P}_{i,j}$  is reduced in Eq. (6.4). To compensate for such values of correlation coefficients, we replace the feasible set given by Eq. (6.4) with the following equation with a tolerance  $\epsilon$ :

$$\widehat{P}_{i,j} = \begin{cases} X : \|X - P_i\|_2 > \|X - P_j\|_2, X \in D, & \text{if } \rho_{j,Query} - \rho_{i,Query} \geq \epsilon, \\ X : \|X - P_i\|_2 \leq \|X - P_j\|_2, X \in D, & \text{if } \rho_{j,Query} - \rho_{i,Query} \leq \epsilon. \end{cases} \quad (6.5)$$

Using the correlation value obtained from all the anchor nodes, the set of feasible points can be reduced further by computing the intersection of all the feasible half-planes as following:

$$\widehat{P}_{Query} = \cap_{i,j} \widehat{P}_{i,j} \quad i, j \in \{1, 2, \dots, K\}, i \neq j. \quad (6.6)$$

As we have ENF data from the five locations, we use four locations as anchor cities and use the ENF data of the fifth city as the query data to estimate its

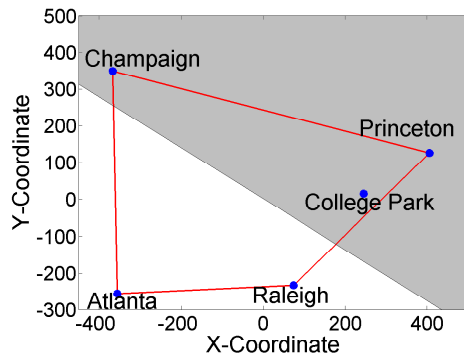


location using the proposed half-plane intersection method. In Figs. 6.9(a)- 6.9(f), we show an example of the set of feasible regions obtained after adding each half-plane constraint, when College Park, Raleigh, Atlanta, and Champaign are the anchor cities and Princeton is the query city. In these figures, the shaded region represents the estimated feasible region for the query city. From this figure, we observe that the area of the feasible region decreases with an increase in the number of constraints, meaning that the precision of localization improves.

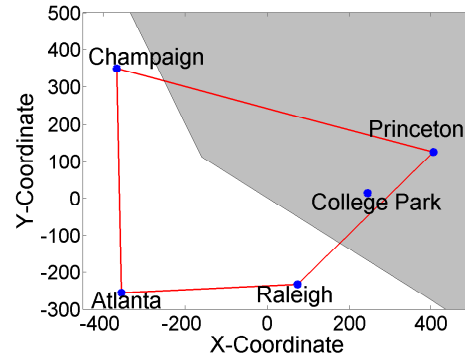
Due to the limited amount of data, we measure the localization performance of the proposed method in terms of two metrics: the probability of localization, denoted by  $p_{loc}$ , and the area of localization, denoted by  $a_{loc}$ . If data from more anchor cities are available, location estimates can be defined using such metrics as the centroid of the feasible set. Among the two metrics we use to evaluate localization performance,  $p_{loc}$  measures the fraction of queries for which the feasible region contains the true location of the query city, i.e.,

$$p_{loc} = \frac{\# \text{ of queries for which } \hat{P}_{Query} \text{ contains } P_{Query}}{\# \text{ of queries}}. \quad (6.7)$$

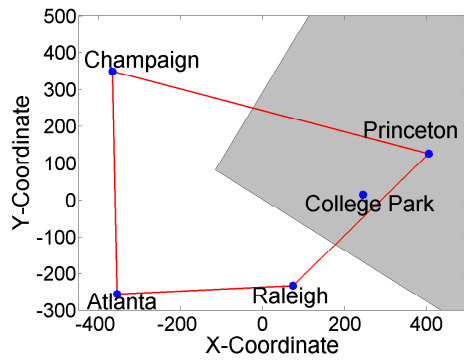
$p_{loc}$  determines the performance of the proposed method in terms of classifying the region of the query city. A higher value of  $p_{loc}$  indicates that the proposed method can localize the query city to a correct feasible region with a high probability. The second performance metric  $a_{loc}$  measures the ratio of the area of the feasible set to the area of the total domain on which the localization is performed, for cases when the feasible set contains the location of the query city. The domain can be defined, for example, as a rectangular region that surrounds all the 5-locations in



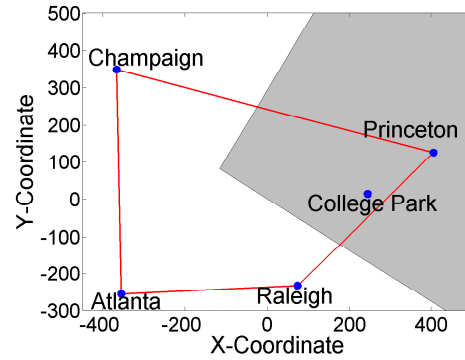
(a) Constraint 1



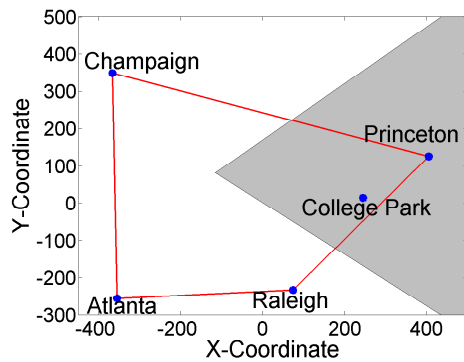
(b) Constraint 2



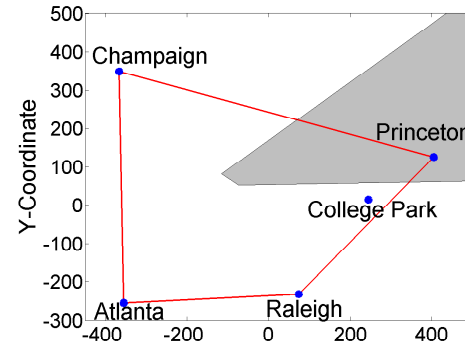
(c) Constraint 3



(d) Constraint 4



(e) Constraint 5



(f) Constraint 6

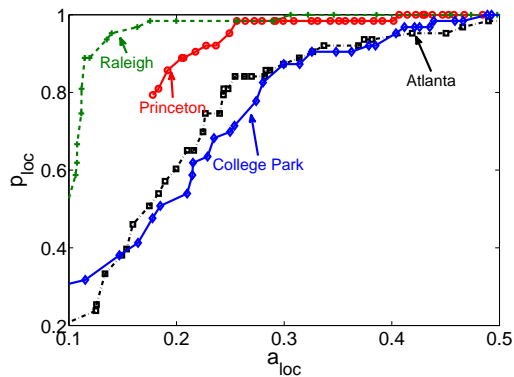
Figure 6.9: Example of localization of Princeton using the half-plane intersection method, when College Park, Raleigh, Atlanta, and Champaign are the anchor nodes. Shaded area represents the feasible region after each additional constraint is applied. The relative positions of the cities with respect to each other are shown.

our dataset.  $a_{loc}$  can be mathematically represented as:

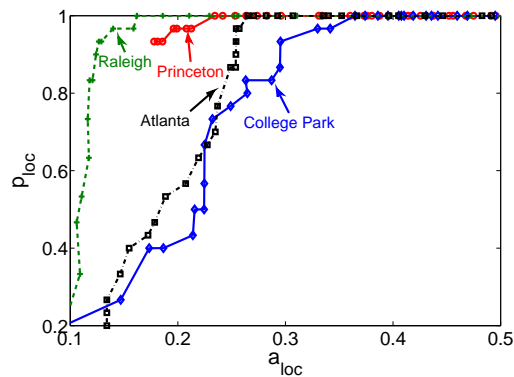
$$a_{loc} = \frac{\text{area of region } \hat{P}_{Query}}{\text{area of domain } D} \cdot \mathbb{1}(P_{Query} \in \hat{P}_{Query}), \quad (6.8)$$

where  $\mathbb{1}(\cdot)$  is an indicator function.  $a_{loc}$  determines the precision of the localization, i.e., smaller the value of  $a_{loc}$ , higher is the localization precision.

Fig. 6.10(a) and 6.10(b) show the plot of  $a_{loc}$  vs.  $p_{loc}$  of different cities by considering other four cities in the dataset as anchor nodes for four-minute long and eight-minute long query segments, respectively. The different values of  $a_{loc}$  and  $p_{loc}$  are obtained by varying the value of  $\epsilon$ . In an ideal scenario, the value of  $p_{loc}$  should be close to one and the value of  $a_{loc}$  should be close to zero. From this plot, we observe that  $p_{loc}$  increases with an increase in the value of  $a_{loc}$  for all the cities (equivalently, the precision decreases). This happens due to the trade-off introduced by the tolerance  $\epsilon$  in  $p_{loc}$  and  $a_{loc}$ . For low values of  $\epsilon$ , the value of  $p_{loc}$  is less, as the hard decision rule does not provide a correct estimate of the feasible area of location of query city, when measurements are noisy. As the value of  $\epsilon$  increase, hard decision boundaries provide a tolerance, which increases the value of  $p_{loc}$ , but a decrement in the number of constraints also reduces the precision. As the query segment duration increases from four to eight minutes, the localization performance also improves as using a large segment size provides a more robust estimate of the correlation coefficient.

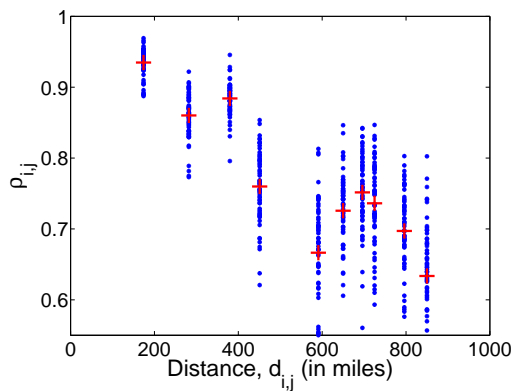


(a) 4-minute query segment

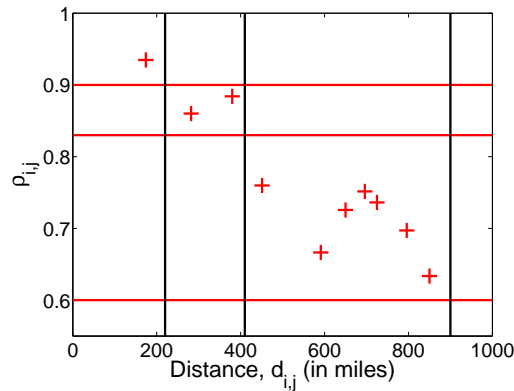


(b) 8-minute query segment

Figure 6.10: Probability of localization,  $p_{loc}$ , and area of localization,  $a_{loc}$  for 5-location US east data using the half-plane intersection method.



(a)  $\rho_{i,j}$  vs  $d_{i,j}$



(b) Quantization of  $\rho_{i,j}$  and corresponding distance bins

Figure 6.11: Relationship between  $\rho_{i,j}$  and  $d_{i,j}$  for 5-location data.

## 6.5 Correlation Quantization for Localization

In this section, we describe a method similar to trilateration [10] for localization using the quantized correlation coefficient information between the city pairs. In this method, we rely on the observation from Figs. 6.7(a)- 6.7(d) that the values of correlation coefficient  $\rho_{i,j}$  between the location signatures of the  $i^{th}$  and the  $j^{th}$  location decrease as the distance  $d_{i,j}$  between them increases. We plot  $\rho_{i,j}$  as a function of  $d_{i,j}$  for the 5-location data in Fig. 6.11(a), which shows that the locations close to each other have a higher value of  $\rho_{i,j}$  as compared for locations further apart. However, it is difficult to derive a functional relationship between the value of  $\rho_{i,j}$  and  $d_{i,j}$ . A close observation of Fig. 6.11(a) reveals that it may be possible to quantize the value of  $\rho_{i,j}$  in a range of distances. One realization of such a quantization is shown in Fig. 6.11(b). In this figure, the red points indicate the mean of the value of  $\rho_{i,j}$  at a particular distance, the blue line indicates the quantization bin on  $\rho_{i,j}$  axis, and the black lines indicate the corresponding distance bins. Based on such a quantization scheme, the following relationship can be derived between  $\rho_{i,j}$  and  $d_{i,j}$ :

$$\left\{ \begin{array}{ll} 100 \leq d_{i,j} < 220, & \text{if } 0.9 < \rho_{i,j} \leq 1 \\ 220 \leq d_{i,j} < 450, & \text{if } 0.83 < \rho_{i,j} \leq 0.9 \\ 450 \leq d_{i,j} < 900, & \text{if } 0.55 < \rho_{i,j} \leq 0.83 \\ 900 \leq d_{i,j}, & \text{if } < \rho_{i,j} \leq 0.55. \end{array} \right. \quad (6.9)$$

Based on the relations in Eq. (6.9), a trilateration based protocol can be derived to estimated the feasible region of recording for a query city. In this trilatera-

tion protocol, the location signatures from an anchor node correlates to a query city to estimate the feasible region by using the constraints from Eq. (6.9) as following:

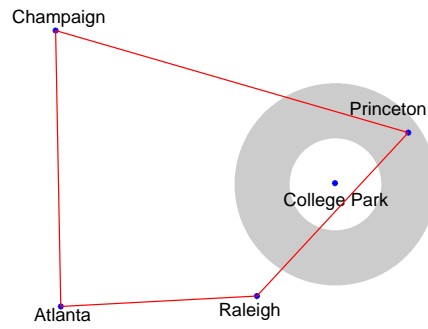
$$\widehat{P}_{i,Query} = \begin{cases} X : 100 \leq \|X - P_i\|_2 < 220, X \in D, & \text{if } 0.9 < \rho_{i,Query} \leq 1 \\ X : 220 \leq \|X - P_i\|_2 < 450, X \in D, & \text{if } 0.83 < \rho_{i,Query} \leq 0.9 \\ X : 450 \leq \|X - P_i\|_2 < 900, X \in D, & \text{if } 0.55 < \rho_{i,Query} \leq 0.83 \\ X : 900 \leq \|X - P_i\|_2, X \in D, & \text{if } \rho_{i,Query} \leq 0.55. \end{cases} \quad (6.10)$$

The feasible region of localization can then be obtained by the intersection of the feasible regions obtained using all the anchor nodes as following:

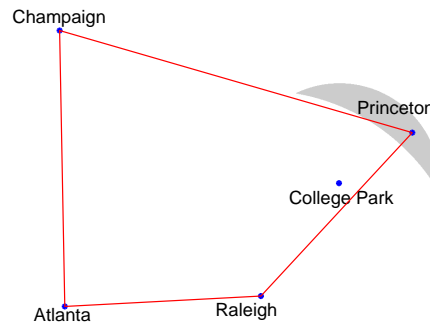
$$\widehat{P}_{Query} = \cap_i \widehat{P}_{i,Query}, i \in \{1, 2, \dots, K\}. \quad (6.11)$$

We measure the performance of the protocol in terms of probability of localization  $p_{loc}$  and area of localization  $a_{loc}$ , as discussed in Section 6.5 and defined in Eq. (6.7) and Eq. (6.8), respectively.

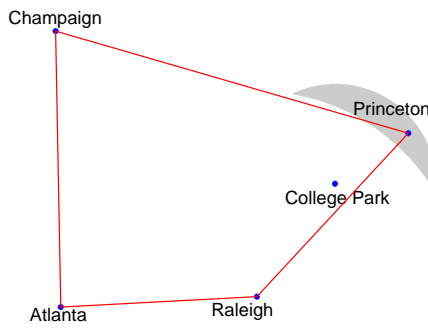
As we have ENF data from five locations, we use four locations as anchor cities and use the ENF data of the fifth city as the query to estimate its location using the correlation quantization relations described in Eq. (6.9) and (6.10). In Figs. 6.12(a)- 6.12(d), we show the set of feasible area obtained after each constraint is added with College Park, Raleigh, Atlanta, and Champaign as anchor cities, and Princeton as the query city. In these figures, the shaded region represents the estimated feasible region for the query city. From this figure, we observe that as the number of constraints increases, the area of the feasible region decreases improving the localization precisions. From Fig. 6.12(d) and Fig. 6.9(f), we observe that the



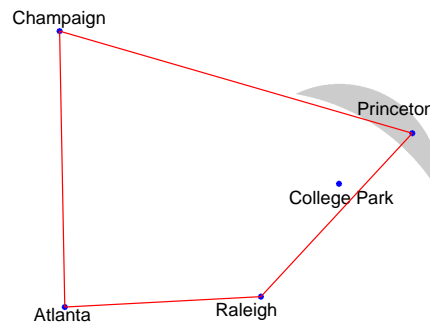
(a) Constraint 1



(b) Constraint 2



(c) Constraint 3



(d) Constraint 4

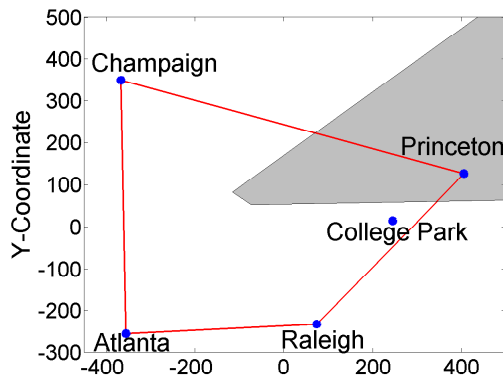
Figure 6.12: Example of localization of Princeton using the correlation quantization method with College Park, Raleigh, Atlanta, and Champaign as anchor locations. Shaded area represents the feasible region after each additional constraint is applied.

localization precision using correlation quantization is better than the half-plane intersection method, as the value of  $a_{loc}$  using the former method is less than the value of  $a_{loc}$  for the latter method.

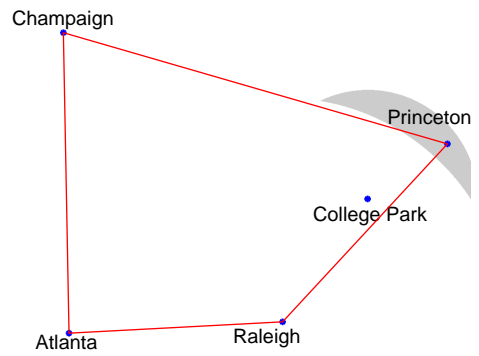
We also combine the localization constraints from the half-plane intersection method in Eq. (6.5) and correlation quantization in Eq. (6.10) to obtain the localization accuracy. In Fig. 6.13(c), we plot an example of the localization feasible area using the combination of half-plane intersection method and correlation quantization method. The corresponding feasible area of localization for the half-plane method and correlation quantization methods, separately, are shown in Figs. 6.13(a) and 6.13(b), respectively. From these figures, we observe that the combination of the two methods improves the area of localization over the half-plane intersection and the correlation quantization.

We plot the localization performance in terms of  $p_{loc}$  and  $a_{loc}$  for all three methods in Figs. 6.14(a)- 6.14(h) for the 5-location data. From these figures, we observe that localization precision for correlation quantization approach is better than the half-plane intersection method for all the cities for higher values of  $\epsilon$ , when the value of  $p_{loc}$  for the half-plane intersection method may be slightly better as compared with the correlation quantization method. Using the constraints from the half-plane intersection method with the correlation quantization method reduced the value of  $p_{loc}$  slightly for a given  $\epsilon$  as compared with using the half-plane intersection method, but improves the localization precision by a significant amount. From these figures, it can be concluded that the use of the combination of the half-plane intersection and the correlation quantization improves the localization performance,

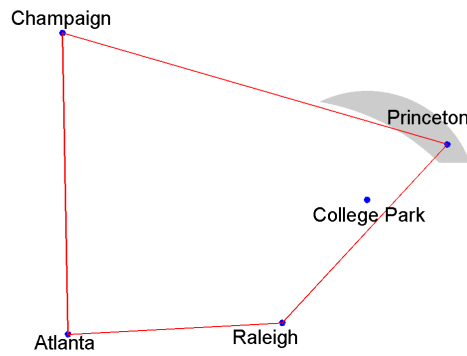




(a) Half-plane method



(b) Correlation quantization method



(c) Half-plane and correlation quantization method

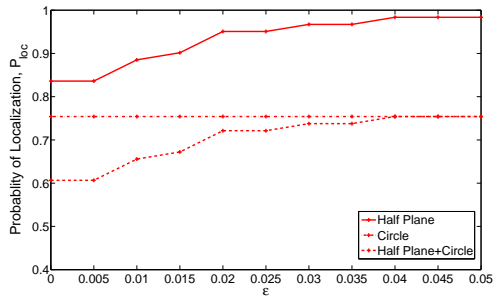
Figure 6.13: Example of localization of Princeton using all the three localization protocols with College Park, Raleigh, Atlanta, and Champaign as anchor locations. Shaded area represents the feasible region after each additional constraint is applied..

as compared to using any method individually.

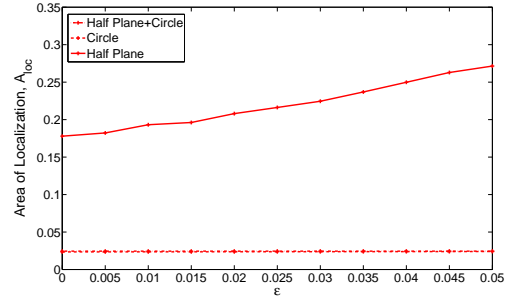
## 6.6 Sensitivity Analysis

In Section 6.3.2 and 6.3.3, we demonstrated the presence of location information signatures in ENF signals. These results are obtained on ENF signals extracted directly from the power mains for the 3-location and the 5-location datasets. ENF signals extracted from power signal are clean signal with very high signal-to-noise ratio. This quality has not been observed in ENF signals embedded in multimedia recordings. It may be possible to use the analysis of Section 6.3.2 and 6.3.3 in multimedia forensics, when the recording device pro-actively captures the ENF signal from power mains and embeds it into the multimedia recording. Such scenarios can be made feasible by adding a power capturing device to audio recorders and cameras, and conducting the recording using the device connected to the power mains. However, in more practical scenarios, it may not be desirable to conduct recordings using a device connected to the power mains.

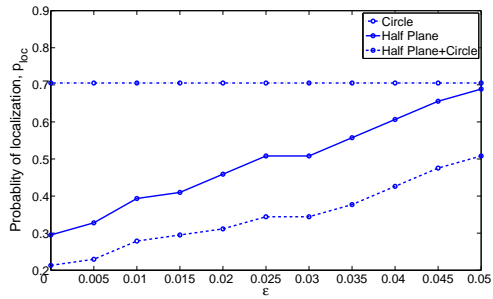
To understand the capabilities and limitations of ENF signal analysis for intra-grid location estimation in multimedia recordings, we conduct a study of the effect of noise on the localization capabilities of ENF signals. We evaluate the localization performance of the half-plane intersection method under the assumption that the ENF signal extracted from the query data is submerged in additive Gaussian white noise. This scenario is representative of cases when ENF signals at anchor nodes are obtained from the power mains recordings, and the query data are multimedia



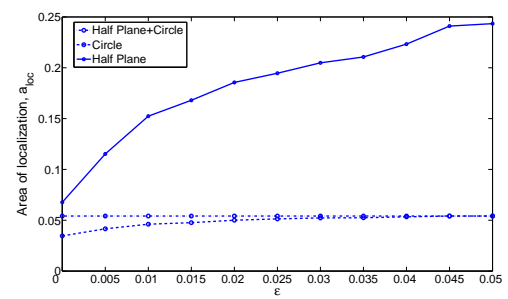
(a)  $p_{loc}$ -Princeton



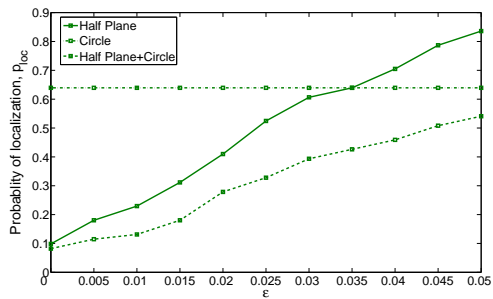
(b)  $a_{loc}$ -Princeton



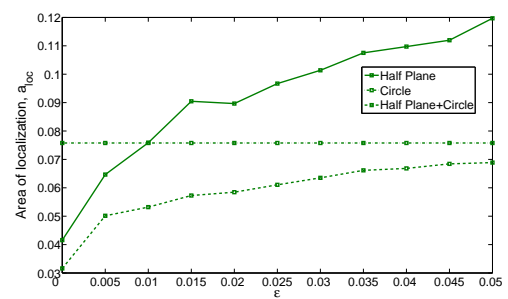
(c)  $p_{loc}$ -College Park



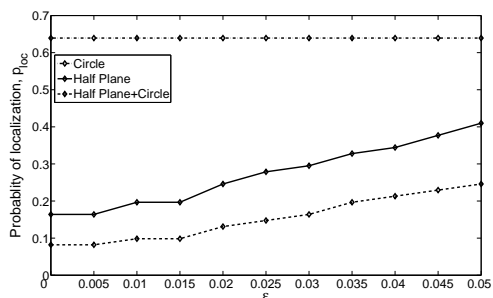
(d)  $a_{loc}$ -College Park



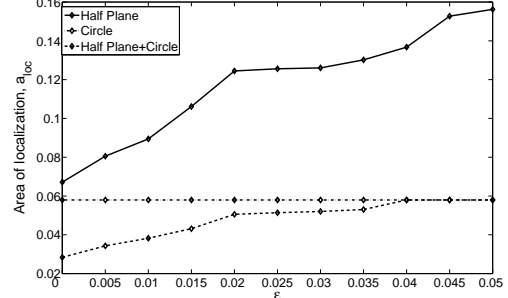
(e)  $p_{loc}$ -Raleigh



(f)  $a_{loc}$ -Raleigh

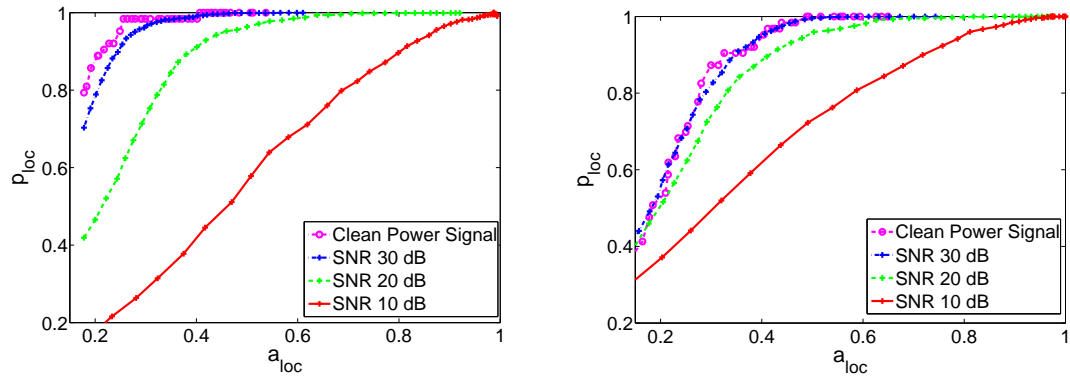


(g)  $p_{loc}$ -Atlanta



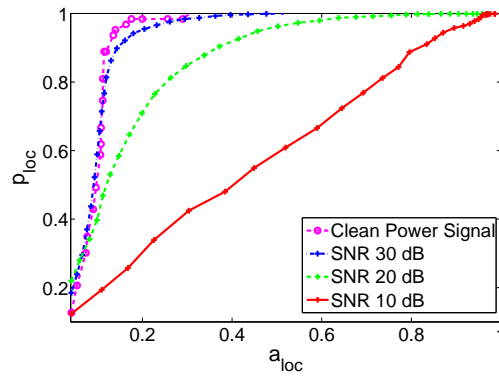
(h)  $a_{loc}$ -Atlanta

Figure 6.14:  $p_{loc}$  and  $a_{loc}$  using different localization methods.



(a) Query city=Princeton

(b) Query city=College Park



(c) Query city=Raleigh

Figure 6.15: Probability of localization  $p_{loc}$  under different noisy conditions for 5-location data using half-plane intersection method.

recordings influenced by ENF signals. We evaluate the effect of different noise levels in the query ENF signal, and plot  $a_{loc}$  v.s.  $p_{loc}$  in Figs. 6.15(a)- 6.15(c) for Princeton, College Park, and Raleigh as query cities, respectively. From these figures, we observe that the localization performance for all three query cities improves with an increase in signal-to-noise ratio (SNR). For  $\text{SNR} > 30$  dB, the localization performance approaches the localization performance obtained using the clean power-ENF signal as the query signal.

As discussed in Chapter 4, the value of SNR observed for the state-of-the-art ENF estimation methods in multimedia signal is in 0-15 dB range. The noise sensitivity analysis clearly indicates that a SNR of more than 30 dB in the query signal is necessary for a good localization performance. Based on these observations, it is difficult to utilize the ENF signals for intra-grid localization in multimedia recordings using the state-of-the-art ENF estimation techniques and localization feature extraction techniques presented in this chapter. Nevertheless, this study demonstrates that intra-grid location specific signatures are present in ENF signals, which can be used to design localization protocols.

## 6.7 Chapter Summary

In this chapter, we have explored location estimation capabilities of ENF signals. For multi-location data, we studied a half-plane intersection method to estimate the location of an unknown recordings. The localization accuracy from this method can be improved by adding more locations as anchor nodes. The number

of constraints to estimate the feasible region increases on the order of  $\mathcal{O}(K^2)$ , with the number of anchor cities  $K$ . We have also demonstrated that a better performance can be obtained by using the constraints based on the proper quantization of correlation coefficient based features. A combination of the half-plane intersection method and the correlation coefficient quantization method provided the best localization performance measured in terms of the probability of localization  $p_{loc}$  and area of localization  $a_{loc}$ .

This chapter focussed on exploring the uncharted application of ENF signal analysis for intra-grid location estimation of multimedia data. The first study conducts experiments on power-ENF signals and provides encouraging results in that direction. Multimedia ENF data are, however, more challenging than power-ENF data because of noise. As we employ the high frequency variations of the ENF signal to extract a meaningful metric for localization, the noisy nature of the ENF signal in multimedia data may increase the difficulty in localization. Furthermore, as shown by our experiments, location specific variations are best captured using instantaneous frequencies estimated at one-second temporal resolution; reliable ENF signal extraction from multimedia data at such a high temporal resolution also presents a research challenge. The results presented in this chapter demonstrate that ENF signals offer a strong potential to be used as a location-stamp.

# Chapter 7

## A Gradient Descent Approach to Secure Localization

### 7.1 Chapter Introduction

In Chapter 6, we have studied the location specific signatures in ENF signals, which can be used to estimate the region of recording within the geographical region covered by the same grid. Based on these signatures, we have derived a distance relationship between the signatures from the recording in question with that from an anchor node. We have proposed a localization method based on half plane intersection, which takes highly quantized location information, and we have improved the precision of the location estimates using more refined quantized information. The reason behind using the quantized information is that it is difficult to derive an explicit relationship between location information and distance between them due to non-homogenous grid density and the insufficient number of location data

available to employ some regression methods.

In this chapter, we explore the problem of localization when an explicit relationship can be determined between location signatures from recordings and distance between their location. For this scenario, such methods as trilateration can be applied to estimate the location of a recording. Trilateration based methods can handle simple measurement noises. However, this type of methods are not robust against a more general setting in which either adversaries are present or there exist sensing equipments that lead to outlier measurements. These adversaries try to inject false location dependent information with an aim to mislead the location estimate. For example, an adversary can modify the ENF signal recording at some anchor nodes in order to modify the distance estimates, which results into an incorrect estimate of location. We address this problem of localization in the presence of malicious adversaries for a general case of sensor network localization. We refer to this problem of localization under potentially adversarial situations as *secure localization*.

The problem of secure localization has been gaining more importance in wireless sensor networks deployed in hostile scenarios such as military surveillance, war-field reconnaissance operations, underwater surveillance [50], forest fire detection [51], flood detection [52], etc. In such applications as forest fire and flood detection [51] [52], sensors alert the base station to any changes in parameters that indicate a potential for forest fire or flooding. The base station needs to know the locations of the nodes transmitting the data, so that appropriate actions may be taken at the relevant site to prevent disasters or to provide early response and contain the damage. In many applications, static sensors may be randomly deployed



within an area using helicopters or road vehicles. As a result, the sensor locations are not known a priori and need to be determined after deployment.

In a hostile environment, an adversary may wish to prevent accurate localization of the nodes and thus prevent the entire network from functioning properly. The adversary may compromise some nodes and thereby gain access to the secret keys and other data stored on the node. This information can then be used to provide misleading information to the base station and other nodes in the network. Incorrect location references may also be provided by intercepting and replaying the packets containing measurements transmitted by anchor nodes. Without effective approaches to filter out or nullify the effect of incorrect measurements, localization would result in a wrong estimate of the sensor position. Hence, there is a strong need to design secure localization algorithms that are robust to such intentional attacks and accurately determine the positions of sensors in the presence of adversaries. At the same time, as the sensors have limited memory, computational, and energy resources, these secure localization algorithms should be resource efficient.

### **7.1.1 Prior Work**

A related problem of location verification has been explored in the literature, where the focus is on developing strategies to verify that a node is indeed located at the claimed position. Methods such as verifiable multilateration, location verification using mobile base stations, and several other distance bounding protocols have been proposed to withstand attacks in secure location verification

problems [53] [54] [55].

The problem of secure localization in WSNs in the presence of malicious adversaries has also attracted attention in the research community. A greedy approach to find the location consistent with the largest number of measurements from anchor nodes was explored in [56]. A voting based scheme was also proposed, in which the localization area is divided into a grid and the vote count of each grid point is incremented if its distance from an anchor node is approximately equal to the distance measurement obtained from that anchor. A similar voting approach with the help of sectored antennas and beacon nodes was proposed in [57]. From a signal processing point of view, the voting based scheme is similar in spirit to the Hough transform used for detecting objects with certain shapes in computer vision and image processing literature [58]. In the Hough transform, a voting procedure is carried out in a parameter space, from which candidate parameters for objects are determined as local maxima of accumulated votes. Similarly, in the voting based scheme for secure localization, the location with the maximum votes is identified as the position of the node.

A Least Median Square (LMdS) approach was proposed in [59] to solve the localization problem for scenarios where less than 50% of the nodes are malicious. This method shares similarities with the Random Sample Consensus (RANSAC) algorithm [60], as it uses several subsets of nodes to identify candidate locations, and then chooses the solution that minimizes the median of the residues. Most of these existing methods localize the nodes with small error as long as the fraction of malicious nodes is not too large. However, the memory requirement and computa-

tional cost of running these algorithms is still high and can be difficult to meet in resource limited applications.

use this information to estimate the location. These prior works assume the presence of some anchor nodes that are used to determine the position of the mobile nodes, and cannot be applied to mobile networks without anchor nodes.

In this chapter, we develop an iterative technique for secure localization with general sensor network applications. In terms of the vector interpretation for iterative updates, the proposed algorithm has similarities to the robust localization algorithm inspired by Self Organizing Maps proposed in [61]. The algorithm in [61] considered noise, but was not designed to withstand attacks by active adversaries, whereas we develop an algorithm for localization that can filter out malicious measurements obtained from nodes compromised by adversaries.

In this chapter, we propose a computationally efficient method to solve the problem of secure localization based on gradient descent. The main idea behind the algorithm is to minimize a suitable cost function involving the position of the localizing node and the available measurements using an iterative gradient descent approach. The cost function is dynamically updated to remove inconsistent measurements arising from malicious nodes. The algorithm operates in two stages. In the first stage, the cost function involves data from all anchor nodes. In the second stage, selective pruning of inconsistent measurements is performed to mitigate the effect of malicious nodes on the solution. The second stage of the algorithm is similar in spirit to the approach employed in [62] to find the Least Trimmed Squares (LTS) solution to data containing outliers [63]. We show that the proposed algorithm can

achieve localization accuracy better than or comparable to existing algorithms in a computationally efficient manner.

## 7.2 Secure Localization in Wireless Sensor Networks

In this section, we consider the secure localization problem for sensor networks where direct measurements of the distance between the localizing node and the anchor nodes are available. These measurements may be obtained through different techniques such as hop count and ToA measurements. When ToA is used to obtain the distance measurements, each anchor node transmits a beacon signal that includes a timestamp and its own location. The localizing node determines its distance from the anchor node based on the embedded timestamp and the time at which it receives the beacon signal. The scenarios where direct measurements of the distance are not available, such as when TDoA is used for localization, will be considered in Section 7.3.

### 7.2.1 Problem Formulation

Let  $N$  be the number of anchor nodes whose locations are known. These may represent nodes that are deployed at known locations and serve to bootstrap the localization of the other sensors in the network. Once a node has determined its own location, it can function as an anchor node for localizing the remaining nodes. Let us

denote the true position of the localizing node by  $\mathbf{P} = [x_{true}, y_{true}]^T$ . The localizing node receives the location of each of the anchor nodes  $\mathbf{P}_k = [x_k, y_k]^T$  and an estimate of the distance between the anchor node and itself, which may be obtained using techniques such as ToA. These distance measurements may be noisy in practice, and we model the measurement errors as additive Gaussian noise with zero mean and variance  $\sigma^2$ . Given the set of noisy measurements  $\{(\mathbf{P}_k, d_k)\}, k = 1, 2, \dots, N$ , an estimate for the node's location  $\hat{\mathbf{P}} = [\hat{x}, \hat{y}]^T$  can be obtained by solving the following over-determined system of equations in a Least Square (LS) sense:

$$\|\mathbf{P}_k - \hat{\mathbf{P}}\| - d_k = 0 \quad k = 1, 2, \dots, N \quad (7.1)$$

Nodes compromised by adversaries may intentionally report wrong information in their  $(\mathbf{P}_k, d_k)$  measurements. In these cases, the LS estimate may be quite far from the true location. Thus, we need secure localization algorithms that are resilient to such attacks.

In the static nodes setting we consider two types of adversaries with different objectives and resources. The first kind of adversaries are able to compromise multiple nodes, but have limited communication and computational resources to coordinate the attacks launched by these nodes. We refer to these attacks as *non-coordinated attacks*. The second type of adversaries have more resources and want to not only prevent the network from precisely locating the nodes, but also try to shift the location estimates to some desired position. We refer to these attacks as *coordinated attacks*. Detailed formulations of these attacks are as follows.

### 7.2.1.1 Non-coordinated Attacks

In non-coordinated attacks, the adversary is assumed to act independently at each compromised node and aims to prevent accurate localization by perturbing the distance estimates reported to the localizing node. Without loss of generality, we assume that each malicious node modifies the  $d_k$  value of the  $(\mathbf{P}_k, d_k)$  measurements, since modifying any other parameter can be transformed into an equivalent modification of the  $d_k$  value. We model the non-coordinated attack by adding independent uniformly distributed perturbations to the actual distance estimates from each malicious node and provide this information to the localizing node. Let  $dist_k = \|\mathbf{P}_k - \mathbf{P}\|$  be the actual distance between the localizing node and the anchor. Define

$$d_k^{(nc)} = \begin{cases} dist_k + u_k + n_k & \text{if node } k \text{ is malicious,} \\ dist_k + n_k & \text{otherwise,} \end{cases} \quad (7.2)$$

where the  $u_k$  are independent zero-mean uniform random variables with variance  $\sigma_{attack}^2$  that model the perturbation introduced from non-coordinated attacks, and the  $n_k$  are independent  $\mathcal{N}(0, \sigma^2)$  Gaussian variables representing the measurement noise. Under the non-coordinated attack setting, the localizing node receives the measurements  $\{(\mathbf{P}_k, d_k^{(nc)})\}, k = 1, 2, \dots, N$  from  $N$  anchor nodes, and uses this information to determine its position.

### 7.2.1.2 Coordinated Attacks

A stronger attack against the network can be launched by multiple compromised nodes acting together to make a localizing node estimate its position as

$\mathbf{P}_{\text{mal}} = [x_{\text{mal}}, y_{\text{mal}}]^T$ , which is some arbitrary point determined by the attackers. We model this scenario by reporting the distance between the anchor node position  $\mathbf{P}_{\mathbf{k}}$  and  $\mathbf{P}_{\text{mal}}$  as the measurement from the malicious anchor. Specifically, let  $d_k^{(c)}$  be defined as:

$$d_k^{(c)} = \begin{cases} \|\mathbf{P}_{\mathbf{k}} - \mathbf{P}_{\text{mal}}\| + n_k & \text{if node } k \text{ is malicious,} \\ \text{dist}_k + n_k & \text{otherwise,} \end{cases} \quad (7.3)$$

where  $\text{dist}_k$  is the actual distance between the  $k$ th anchor and the localizing node and  $n_k$  represents measurement noise as before. The localizing node receives the measurements  $\{(\mathbf{P}_{\mathbf{k}}, d_k^{(c)})\}, k = 1, 2, \dots, N$  from the anchor nodes and uses this information to determine its position. The strength of the coordinated attack is characterized in terms of the distance,  $d_a = \|\mathbf{P}_{\text{mal}} - \mathbf{P}\|$ , between the actual position and the position reported by malicious nodes.

## 7.2.2 Proposed Method for Secure Localization

In this subsection, we propose an iterative secure localization algorithm by combining gradient descent with a selection stage to filter out the malicious measurements [64]. We first consider the likelihood of the measurements given the true position of the localizing node. When there is no malicious node, and the measurement noise is Gaussian, the likelihood of the measurements given the true position of the localizing node  $\mathbf{P}$  is

$$\Pr(\{d_k\}_{k=1}^N | \mathbf{P}, \{\mathbf{P}_{\mathbf{k}}\}_{k=1}^N) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^N (\|\mathbf{P}_{\mathbf{k}} - \mathbf{P}\| - d_k)^2\right\}. \quad (7.4)$$

The Maximum Likelihood (ML) estimate  $\hat{\mathbf{P}}$  for the true position can then be found by maximizing the likelihood of the measurements, or equivalently, by minimizing the negative of the exponent:

$$\begin{aligned}
\hat{\mathbf{P}} &= \arg \max_{\mathbf{P}} \Pr (\{d_k\}_{k=1}^N | \mathbf{P}, \{\mathbf{P}_k\}_{k=1}^N) \\
&= \arg \min_{\mathbf{P}} \frac{1}{2} \sum_{k=1}^N (\|\mathbf{P}_k - \mathbf{P}\| - d_k)^2 \\
&= \arg \min_{\mathbf{P}} f(\mathbf{P}),
\end{aligned} \tag{7.5}$$

where  $f(\mathbf{P})$  denotes the cost function in Eq. (7.5) and corresponds to the negative of the exponent in Eq. (7.4). The ML estimate  $\hat{\mathbf{P}}$  is thus identical to the LS estimate obtained by solving Eq. (7.1) in a least squares sense.

In the proposed secure localization algorithm, we adopt an iterative gradient descent algorithm to first search for the LS solution. The algorithm starts by randomly initializing the estimate  $\hat{\mathbf{P}}(0)$  to some point in the deployment area. At the  $i^{\text{th}}$  step of the iteration, the gradient of the cost function  $f(\mathbf{P})$  is evaluated at the current estimate  $\hat{\mathbf{P}}(i-1)$ , and the estimate is then updated by moving it one step in the direction of the negative of the gradient. Let  $\mathbf{g}(i)$  denote the negative of the gradient of the cost function at the current estimate of the position:

$$\mathbf{g}(i) = - \nabla_{\mathbf{P}} (f(\mathbf{P}))|_{\mathbf{P}=\hat{\mathbf{P}}(i-1)}, \tag{7.6}$$

where  $\nabla_{\mathbf{P}}(\cdot)$  denotes the derivative with respect to  $\mathbf{P}$ . The estimate is then updated by moving it one step in the direction of the negative of the gradient as

$$\hat{\mathbf{P}}(i) = \hat{\mathbf{P}}(i-1) + \delta(i) \times \frac{\mathbf{g}(i)}{\|\mathbf{g}(i)\|}, \tag{7.7}$$



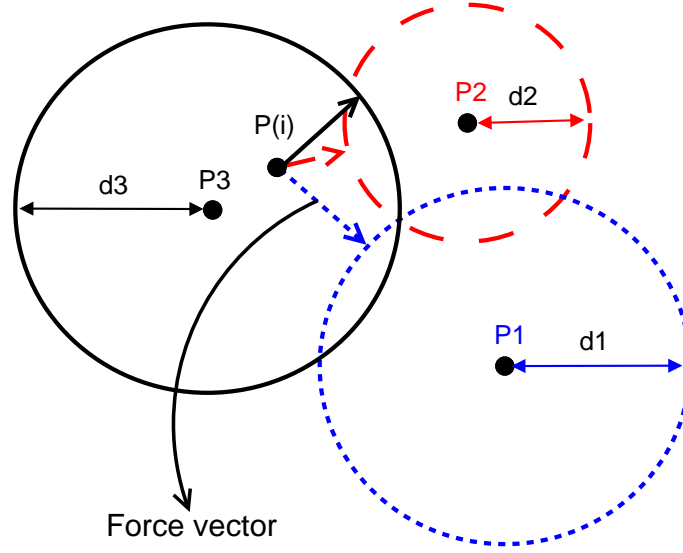


Figure 7.1: Force Vector representation of terms contributing to the gradient.

where  $\delta(i)$  is the step size at the  $i^{th}$  iteration and  $\frac{\mathbf{g}(i)}{\|\mathbf{g}(i)\|}$  is the unit vector in the direction of the negative of the gradient. The negative gradient  $\mathbf{g}(i)$  is found to be:

$$\begin{aligned}
 \mathbf{g}(i) &= -\nabla_{\mathbf{P}} (f(\mathbf{P}))|_{\mathbf{P}=\hat{\mathbf{P}}(i-1)} \\
 &= -\nabla_{\mathbf{P}} \left( \frac{1}{2} \sum_{k=1}^N (\|\mathbf{P}_{\mathbf{k}} - \mathbf{P}\| - d_k)^2 \right) \Big|_{\mathbf{P}=\hat{\mathbf{P}}(i-1)} \\
 &= \sum_{k=1}^N (\|\mathbf{P}_{\mathbf{k}} - \hat{\mathbf{P}}(i-1)\| - d_k) \times \frac{\mathbf{P}_{\mathbf{k}} - \hat{\mathbf{P}}(i-1)}{\|\mathbf{P}_{\mathbf{k}} - \hat{\mathbf{P}}(i-1)\|} \\
 &= \sum_{k=1}^N \mathbf{g}_{\mathbf{k}}(i), \tag{7.8}
 \end{aligned}$$

where we define the term  $\mathbf{g}_{\mathbf{k}}(i)$  as:

$$\mathbf{g}_{\mathbf{k}}(i) = (\|\mathbf{P}_{\mathbf{k}} - \hat{\mathbf{P}}(i-1)\| - d_k) \times \frac{\mathbf{P}_{\mathbf{k}} - \hat{\mathbf{P}}(i-1)}{\|\mathbf{P}_{\mathbf{k}} - \hat{\mathbf{P}}(i-1)\|}. \tag{7.9}$$

Conceptually, as shown in Fig. 7.1, the gradient component  $\mathbf{g}_{\mathbf{k}}(i)$  can be visualized as a “force vector” with direction along the line joining the current estimate of the location  $\hat{\mathbf{P}}(i-1)$  and the position of the anchor node  $\mathbf{P}_{\mathbf{k}}$  and magnitude equal to the distance between the current estimate and the circle of radius  $d_k$  around the

anchor node. The sum of these force vectors gives the overall gradient.

Each iteration results in a new estimate that has a higher probability of being the true location of the node. This gradient descent algorithm eventually converges to the ML estimate which is the same as the LS estimate when in the absence of the malicious nodes. As described previously, due to malicious measurements from adversaries, the LS estimate can have large errors. Hence, once the gradient descent algorithm converges to the LS solution, we switch to a selection stage in which some force vectors are pruned as discussed next.

**Selection Stage:** In the non-coordinated attack case, the independent perturbations added by various malicious nodes tends to average out and the LS solution is close to the true position. In the coordinated attack case whereby less than 50% of the nodes are malicious, the LS estimate obtained from the first stage of the algorithm is closer to the true position  $\mathbf{P}$  than to the position  $\mathbf{P}_{\text{mal}}$  chosen by the malicious nodes. This is because in such situations, the true position satisfies more equations in Eq. (7.1) than the position reported by the malicious nodes. So the LS solution tends to be closer to the true position than  $\mathbf{P}_{\text{mal}}$ .

As a result, when the estimate of the node's position  $\hat{\mathbf{P}}(i)$  approaches the LS estimate, the residues corresponding to the terms arising from the malicious nodes tend to be larger than those from the honest nodes. We update the cost function to exclude the terms with large residues, which are likely to correspond to the measurements from the malicious nodes. In our algorithm, we achieve this by pruning out a fraction of the force vectors with large magnitudes and using the remaining vectors to compute the gradient. The estimate is updated by moving

Table 7.1: Comparison of run time complexity of different localization algorithms.

Method	Complexity	Run time(in ms)
Least Median Square	$\Theta(M_1N)$	24.8
Voting based scheme	$\Theta(n_1^2N)$	14.8
Gradient Descent	$\Theta(MN)$	3.7

( $N$  is the number of anchor nodes,  $n_1$  is the size of the grid in the voting scheme, and  $M$  is the number of iterations for the gradient descent algorithm. The values of the parameters are described in Section 7.2.4.)

it one step in the direction of this modified gradient at each iteration. The final algorithm is shown in Algorithm 1, and Fig. 7.2 shows a flowchart of the proposed algorithm.

### 7.2.3 Comparison of Computational Complexity

We now compare the computational complexity of the proposed gradient descent algorithm to the voting scheme [56] and the LMdS scheme [59]. Table 7.1 shows the computational complexity for each algorithm and the average run time for a set of experiments conducted on MATLAB platform. From this table, we see that for the voting scheme, the complexity increases with the square of the grid size. To obtain better localization accuracy, the grid needs to be quantized more finely, leading to a higher number of cells in the grid. Alternatively, grids can be coarsely quantized in the beginning and the localization resolution can be improved by conducting multiple stages of the voting algorithm, with each stage using progressively finer cells in the areas of the grid which receive high number of votes in the previous stage. This leads to high computational requirements in the voting scheme.

---

**Algorithm 1:** Proposed algorithm for secure localization

---

**Input:**  $N$ =number of anchor nodes;  $M$ = number of iterations;  $\delta(\cdot)$ =step size

function (constant or non-increasing)  $S$ ={all anchor nodes};

$\{(\mathbf{P}_k, d_k)\}$ , where  $\mathbf{P}_k = [x_k, y_k]^T$ ,  $k = 1, 2, \dots, N$  be the position

claimed by  $k^{th}$  anchor node, and  $d_k$  be the distance reported to the

localizing node.

**Output:** Estimated coordinates  $\hat{\mathbf{P}}(M) = [\hat{x}, \hat{y}]^T$

**Initialization:** a random point  $\hat{\mathbf{P}}(0)$ ;  $stage = 1$

**for**  $i = 1 : M$  **do**

**for**  $k = 1 : N$  **do**

$$\mathbf{g}_k(i) = (\|\mathbf{P}_k - \hat{\mathbf{P}}(i-1)\| - d_k) \times \frac{\mathbf{P}_k - \hat{\mathbf{P}}(i-1)}{\|\mathbf{P}_k - \hat{\mathbf{P}}(i-1)\|};$$

**end**

$$\mathbf{g}(i) = \sum_{k=1}^N \mathbf{g}_k(i); \quad // \text{gradient}$$

**if**  $(\|\mathbf{g}(i)\| < \text{threshold})$  or  $(stage == 2)$  **then**

$stage = 2$ ;     //enter the selection stage

$S = \{\text{set of } \frac{N}{2} \text{ anchor nodes whose corresponding force vectors are the smallest}\}$ ;

**end**

$$\mathbf{g}(i) = \sum_{k \in S} \mathbf{g}_k(i); \quad // \text{update gradient}$$

$$\text{Update: } \hat{\mathbf{P}}(i) = \hat{\mathbf{P}}(i-1) + \delta(i) \times \frac{\mathbf{g}(i)}{\|\mathbf{g}(i)\|};$$

**end**

---

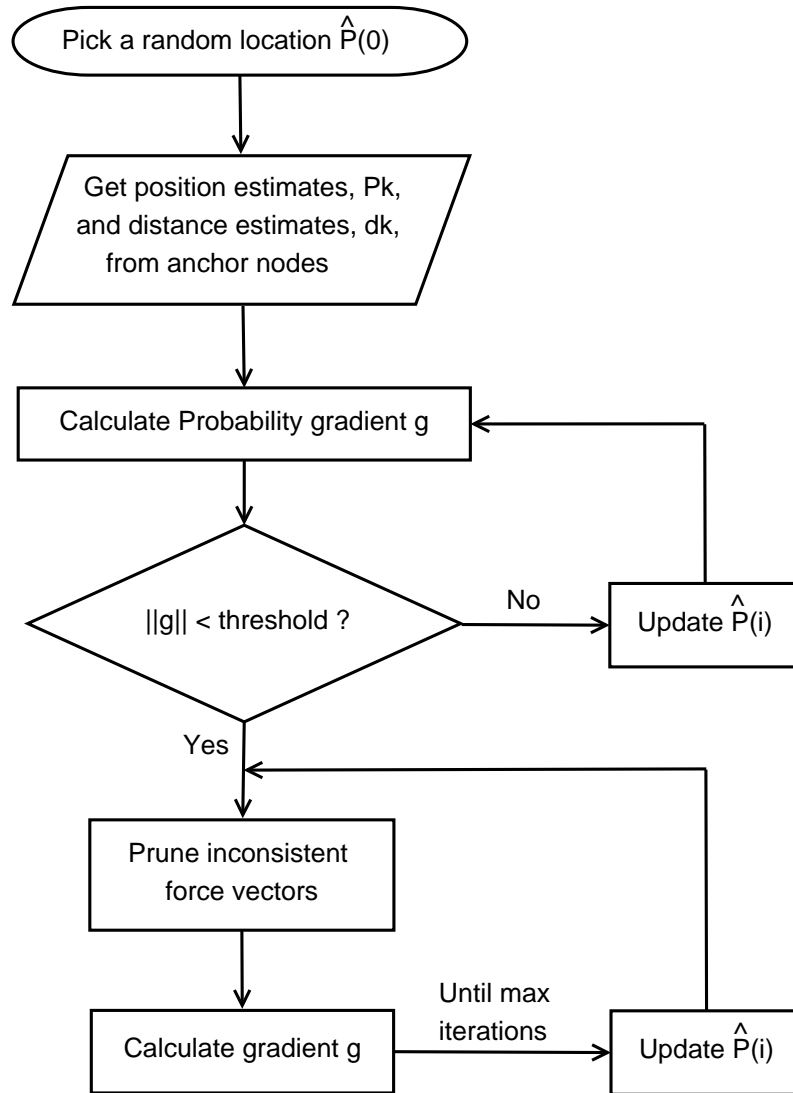


Figure 7.2: Flow chart for localization algorithm.

The LMdS approach requires a certain minimum number of subsets of nodes  $M_1$ , which increases as the percentage of malicious nodes increases, in order to ensure that one estimate is the correct estimate with very high probability. An LS estimate needs to be found for each of these subsets, which is computationally expensive. The computation complexity associated with the LMdS method is calculated using the Linear Least Squares (LLS) algorithm described in [65]. LMdS algorithm first performs  $M_1$  LLS on different subsets of size  $n$  giving a computational complexity of  $\Theta(M_1 n)$ . After finding each LLS solution, a consistency check with measurements from all  $N$  nodes is performed, which has a computational complexity of  $\Theta(M_1 N)$  for all  $M_1$  rounds. In the final step, another LS estimate is found using the maximum size subset of nodes that have passed the consistency test. The computation complexity of this final operation is smaller than that of the first two steps. The overall computation complexity of LMdS is  $\Theta(M_1(N+n))$ , which can be represented as  $\Theta(M_1 N)$ , due to  $n < N$ .

In contrast, the computational complexity of the proposed scheme is independent of the number of malicious nodes and the grid size, although it increases linearly with the number of iterations. The number of iterations can be reduced by choosing variable step size to increase the convergence rate of the algorithm [66]. At each iteration, our proposed algorithm calculates only the distance of the current estimate from the anchor nodes and requires less computation. Thus, the gradient descent algorithm is computationally simpler than the voting based scheme and the LMdS method.

In Fig. 7.3, we plot the run time required to achieve a desired localization

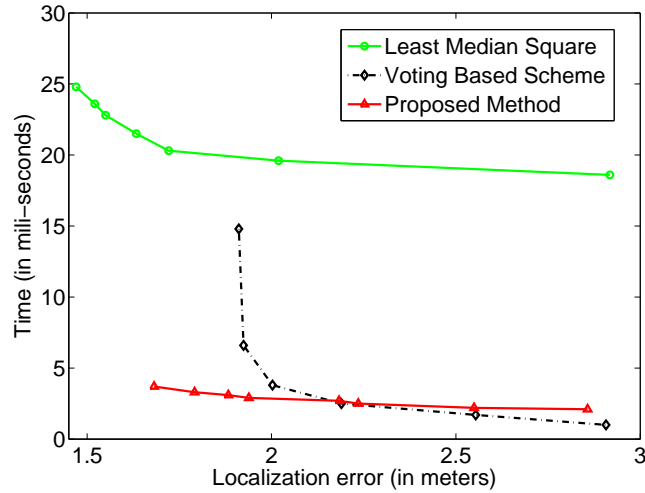


Figure 7.3: Comparison of run-time for different localization schemes for a fixed localization error.

accuracy for different secure localization algorithms considered in this paper. From this plot, we see that the run time required to achieve a given localization accuracy for our proposed method is approximately 8 times lower as compared to the LMdS method. Comparing the voting scheme and gradient descent based scheme, we can see that they have similar run time at settings that result in medium to high location error; Location accuracy for the voting scheme can be improved by increasing the grid density, which leads to a quadratic increase of run time and memory use. The accuracy for our proposed gradient descent scheme can be controlled by the terminating condition of the iteration, and the run time for higher location accuracy increases quite moderately and is considerably lower than the other two schemes.

## 7.2.4 Simulation Results

We experimentally compare our proposed algorithm with two existing secure localization methods, namely, the voting scheme and the LMdS algorithm. The simulation parameters are similar to those in [56] to allow for comparison of the results. 30 anchor nodes are randomly deployed in an area of size  $60\text{m} \times 60\text{m}$ . The measurement noise standard deviation is set to be  $\sigma = 2\text{m}$ . For the LMdS method, the number of subsets is set to be  $M_1 = 20$  and the number of nodes in each subset is chosen to be  $n = 4$ . For the voting scheme, the region of deployment is divided into a square grid with each cell of size  $1\text{m} \times 1\text{m}$ , so that  $n_1 = 60$ . We use the algorithm described in [56] to find the votes for each cell. For the proposed algorithm, in the selection stage, we prune 50% of the force vectors with the largest magnitude and the number of iterations  $M = 200$ . The threshold for switching to selection stage is determined experimentally by varying its value between 0.01 and 0.1 and choosing the value that gives the best localization performance. For our simulations, we determine the threshold value to be 0.9. The results shown are obtained by averaging over 1500 runs of simulations.

We compare the performance of our proposed method when a variable step size and a fixed step size are used for the gradient descent based method, respectively. In the fixed step size version of the algorithm,  $\delta(i) = 0.5$  for all iterations  $i = 1, 2, \dots, M$ . For the variable step size algorithm, we adopt the following step size that is linearly decreasing:



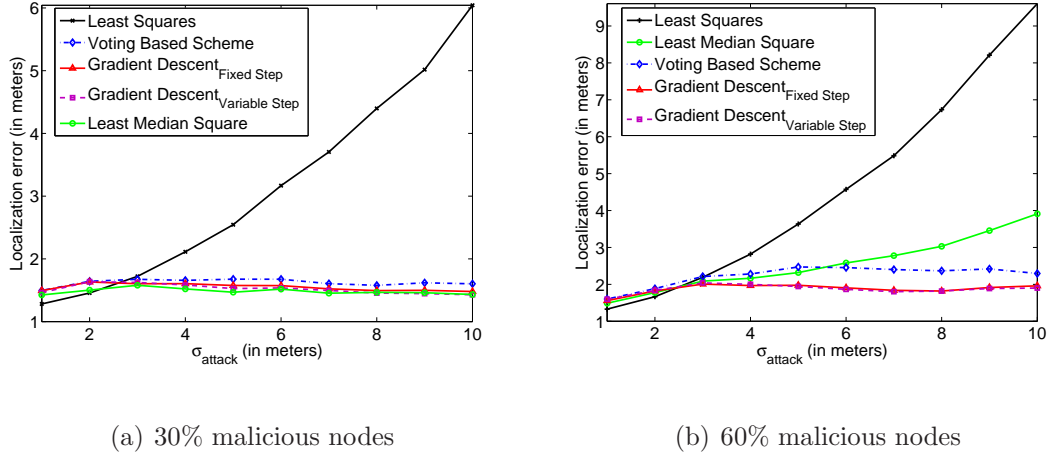


Figure 7.4: Comparison of localization schemes for non-coordinated attacks.

$$\delta(i) = 15 - \frac{15(i-1)}{M} \quad (7.10)$$

### 7.2.4.1 Non-coordinated attacks

The localization accuracy achieved by various secure localization algorithms under non-coordinated attacks with different parameters is shown in Fig. 7.4. In particular, Fig. 7.4(a) shows the localization error as a function of the noise standard deviation  $\sigma_{\text{attack}}$  added by the malicious nodes when 30% of the nodes are compromised; and Fig. 7.4(b) shows the corresponding results when 60% of the nodes are compromised.

From Fig. 7.4(a) we observe that the localization error using our method is comparable to the other schemes when the fraction of malicious nodes is less than 50%. For 60% malicious nodes, the LMdS method results in a localization error that increases with attack strength, as it cannot tolerate attacks by more than 50% of the nodes, but the proposed method can still localize the node with high accu-

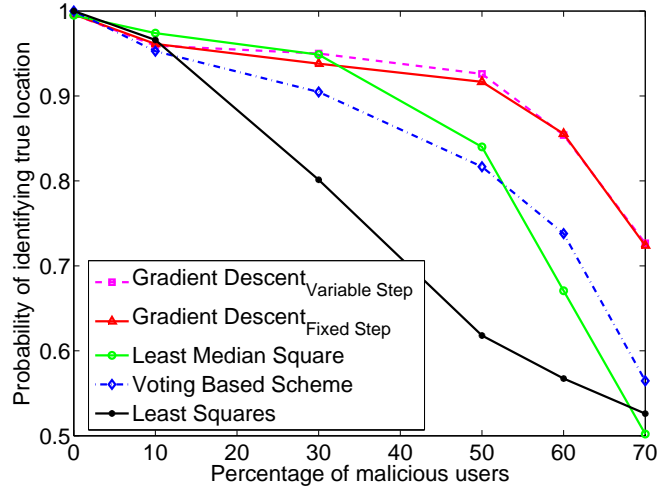


Figure 7.5: Probability of converging to the correct estimate for different localization schemes under non-coordinated attacks for  $\sigma_{attack} = 4m$ .

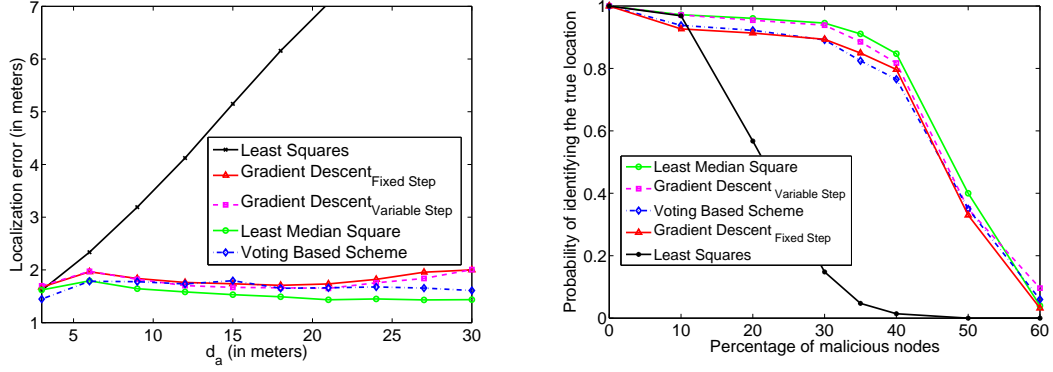
racy. The independent distance random perturbations by the malicious nodes result in randomly oriented force vectors, which have a mutually canceling effect when summed up to compute the overall gradient. As a result, the proposed algorithm is robust against non-coordinated attacks, and the average localization error does not increase as the attack noise variance  $\sigma_{attack}^2$  increases. The voting based scheme also gives good localization accuracy, but the error is slightly higher than the gradient descent method because of the discrete nature of the grid points. The localization accuracy in the voting based scheme can be increased by finely quantizing the grid points at a cost of higher computation and memory.

While the average localization error is a useful indicator of the accuracy of the algorithm, it may be dominated by the cases where the algorithm does not converge to the true position. To obtain a different perspective on the accuracy, we compare

the probability that the algorithm correctly identifies the true position. Due to the presence of noise, finite grid size, and step size, we consider that the algorithm has converged to the correct location if the final estimate is within a distance of  $\sigma$  meters from the true location. Fig. 7.5 compares the probability of converging to the correct estimate as a function of the fraction of malicious nodes participating in the attack, for different localization schemes under non-coordinated attacks. We see that when the fraction of malicious nodes is less than 50%, all the secure localization algorithms except the simple LS method have similar performance and converge to the correct estimate about 90% of the time. However, if the fraction of nodes participating in the attack is more than 50%, the proposed scheme outperforms the existing algorithms. For example, when the fraction of attacking nodes is 60% or 70%, the gradient descent approach has approximately 10% higher probability of converging to the true position.

#### 7.2.4.2 Coordinated attacks

Fig. 7.6(a) shows the localization error under coordinated attack by 30% of the nodes. The x-axis represents the distance  $d_a$  between the true location of the sensor and the point  $\mathbf{P}_{\text{mal}}$  chosen by the malicious nodes at random. From the figure, we observe that when the fraction of malicious nodes is 30%, the localization accuracy for all the methods except LS is almost the same. We obtained similar results when the fraction of malicious nodes is 35%. The localization error for the proposed gradient descent method is slightly higher than the other techniques under this setting. The reason for this behavior is that as the percentage of malicious nodes



(a) Average localization error

(b) Probability of correct localization

Figure 7.6: Performance of the secure localization schemes under coordinated attacks by 30% of the nodes: (a) Average localization error and (b) Probability of correctly identifying the true position for  $d_a = 22m$ .

increases, even a few uncompromised anchor nodes whose distances from malicious position and true position are approximately the same can cause received data from anchor nodes to be more consistent with malicious positions. This phenomenon will be discussed in detail in Section 7.2.5.

Fig. 7.6(b) compares the probability of converging to the true location for various algorithms under coordinated attacks when  $d_a = 22m$ . From this figure, we see that all the secure localization schemes have similar performance and converge to the correct estimate about 90% of the time for coordinated attacks by less than 30% of the nodes. For attacks by a larger fraction of nodes, the probability of converging to the true position is slightly lower for the gradient descent algorithm with variable step size when compared to the LMdS algorithm. This is again due to the honest nodes that are at approximately the same distance from both the true position and

the position reported by the malicious nodes.

Measurement noise also has a major impact on the performance of localization algorithms. In our simulations, we observed that the localization error is approximately same for a fixed  $\sigma$ , and increases linearly with an increase in  $\sigma$  for all three secure localization methods considered in this paper.

### 7.2.5 Discussions

In the first stage of our algorithm, we find the LS estimate of the location in an iterative manner. After convergence in the first stage, our algorithm switches to the second stage to prune outliers. Modeling the secure localization problem in such an iterative framework helps us see similarities between our proposed algorithm and the iterative LTS algorithm, and understand the robustness of our proposed algorithm. In particular, our pruning stage can be modeled similar to the iterative approach used to solve the LTS problem proposed in the literature for robust estimation of the parameters of observations containing outliers [62].

To demonstrate this similarity, each term inside the summation in Eq. (7.5) can be considered as the residual error in estimating the true location  $\mathbf{P}$ . We can rewrite Eq. (7.5) in the following form,

$$\hat{\mathbf{P}} = \arg \min_{\mathbf{P}} \sum_{k=1}^N r_k^2 \quad (7.11)$$

where  $r_k = (\|\mathbf{P}_k - \mathbf{P}\| - d_k)$  denotes the residual in estimating the true location. Let  $(r^2)_{1:n} \leq (r^2)_{2:n} \leq \dots \leq (r^2)_{n:n}$  be the ordered squared residuals of the set

$\{r_1, r_2, \dots, r_n\}$ . The LTS method seeks to minimize the following cost function,

$$\sum_{i=1}^h (r^2)_{i:n} \quad (7.12)$$

where  $h$  is the number of residues used to evaluate the LTS cost function. An efficient iterative method to solve the LTS problem was proposed in [62]. In this iterative approach, parameters estimated in the  $(i - 1)^{th}$  iteration are used to calculate the residues,  $r_k$ 's in the  $i^{th}$  iteration. The residues are then arranged in an ascending order of their magnitudes and an estimate of parameters is obtained for the  $i^{th}$  iteration by finding the LS estimate using the  $h$  smallest residue points. Our proposed method is similar to this general statistics approach, and the magnitude of our force vectors,  $\mathbf{g}_k(i), i = 1, 2, \dots, n$ , reflect the magnitude of the residues,  $r_k, k = 1, 2, \dots, n$ . However, in the iterative LTS method, an LS estimate is obtained at each iteration, which requires rather high computation power. Instead of following the conventional iterative LTS method [62], we iteratively update the location estimate by a step size in the direction of the decreasing cost function. We see from experiments that our algorithm converges even after relaxing the parameter update criteria of the conventional iterative LTS algorithm.

The breakdown point of the iterative LTS algorithm, i.e., the number of outliers that the algorithm is guaranteed to tolerate, is shown in the literature to be approximately 50%, which is the same as that of the LMdS method. As the proposed algorithm shares the spirit of the iterative LTS algorithm, the proposed algorithm has the same breakdown point. However, in non-coordinated attack case, our algorithm can tolerate more than 50% malicious nodes because of the statistically

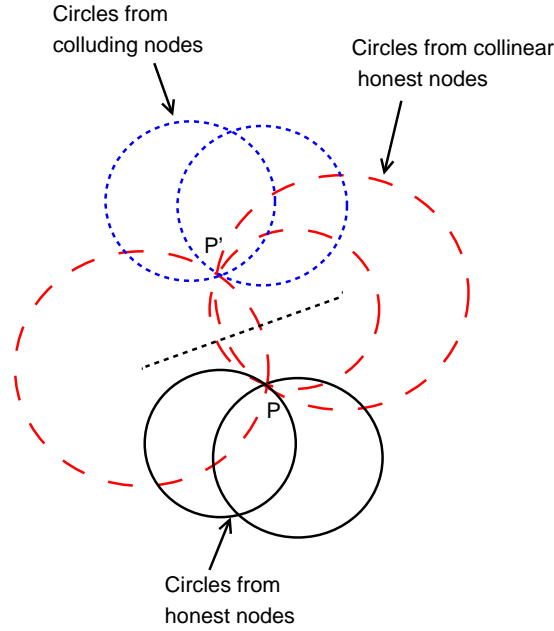


Figure 7.7: Geometry for the bound on the maximum number of colluding nodes.

canceling nature of the attacks as discussed in Section 7.2.4.1. In coordinated attacks, depending on the topology of the sensor nodes, the breakdown point can be slightly less than 50% as discussed next.

Fundamentally under coordinated attacks, it is impossible for any scheme to perform secure localization using only the location and the distance information if the number of coordinating malicious nodes is more than the number of honest nodes. In these cases, there are more consistent equations satisfied by the location ( $\mathbf{P}_{\text{mal}}$ ) reported by the malicious nodes than those satisfied by the true node location ( $\mathbf{P}$ ). Hence, without any additional information to authenticate, it is impossible to distinguish between these two locations, and robustness against coordinated attacks should be focused on the situation when less than 50% of the nodes are compromised. This scenario was analyzed in [67] and it was shown that if the total number of nodes,  $N$ , in the network is more than  $2L + 2$ , where  $L$  is the number of malicious nodes,

then the position of the localizing node can be computed with a bounded error.

In practical scenarios, adversaries can create more consistent measurements for the malicious location even when they are unable to compromise a majority of the nodes, by carefully choosing the distance measurement to report, as shown in Fig. 7.7. In this figure,  $\mathbf{P}$  denotes the location of the localizing node, while  $\mathbf{P}'$  denotes the position chosen by malicious nodes to shift the estimate. Each circle is drawn with position reported by anchor nodes as center and distance measured by localizing node from corresponding anchor node as radius. The circles intersecting at both locations  $\mathbf{P}$  and  $\mathbf{P}'$  correspond to the collinear anchor nodes as centers and are shown using the dashed line. The colluding nodes can take advantage of the fact that the measurements reported by the collinear honest nodes are also consistent with their second point of intersection  $\mathbf{P}'$ . When nodes are randomly deployed, the probability that three or more nodes are exactly collinear is negligible. As such, in the absence of measurement noise, we then require that  $N > 2L + 2$  for secure localization. In practice, however, the presence of measurement noise requires a localization algorithm to equip with some tolerance capability. Therefore, being merely close to collinear in the noisy case can have the inevitable effect to aid the adversary in the same way as what the exact collinear situation does for the noise-free case. Since the probability of nodes being close to collinear is non-trivial, the required lower bound on  $N$  is considerably larger than  $2L + 2$ . In our experiments, we have observed that such occurrences of almost collinear nodes with an intersection point close to  $\mathbf{P}_{\text{mal}}$  account for a large fraction of the cases where the secure localization algorithms do not correctly identify the true position of the



node. Because of increase in the probability of nodes that are collinear, the proposed scheme can tolerate fewer malicious nodes - about 40% in our study as shown in Fig. 7.6(b), for coordinated attack case. Beyond that, the probability of correctly identifying the location drops sharply.

In LMdS, localization is performed using multiple subsets of four nodes and the estimated location will be correct as long as one of these  $M_1$  subsets contains all innocent nodes. Therefore, the LMdS algorithm performs moderately better than the gradient descent and voting algorithms when the percentage of malicious nodes is higher, although this gain is achieved at the cost of much higher computational resources. The difference in localization error between the voting based scheme and our scheme can be attributed to resolution accuracy of grid used in voting scheme and to the step size in the gradient descent based scheme. Overall, we see a tradeoff between the localization accuracy and resource use. Our gradient descent based algorithm requires a significantly less amount of computational and memory resources, at a cost of slightly higher localization error for coordinated attacks launched by a high percentage of colluding malicious nodes.

## 7.3 Gradient Descent Approach Applied to TDoA Measurements

In the previous section, we showed that the proposed gradient descent algorithm with selective pruning can be used to securely localize nodes in hostile

scenarios. We assumed that a direct measurement of the distance between the localizing node and the anchor nodes is available. This distance measurement may be obtained using ToA of the beacon signals, and requires synchronization between the transmitter and the receiver. A small synchronization error can cause a large error in the spatial localization as time is multiplied by the speed of light or sound. Time difference of Arrival (TDoA) is used as one way to mitigate these synchronization issues [68] [69].

The setting for obtaining one TDoA measurement is shown in Fig. 7.8. In this example, the localizing node wants to obtain a TDoA measurement with the help of anchor nodes 1 and 2, which already know their position. Node 2 transmits its position coordinates and a timestamp to node 1 and the localizing node. Node 1 receives the signal from node 2 and forwards it to the localizing node after including node 1's own position coordinates. The forwarding delay in this process is assumed to be known in advance, as it depends on the processing speed at the node and may be known a priori. In order to take account of possible minor variations, we model the forwarding delay to be normally distributed around a known mean value. In practical applications, additional delays associated with queueing may be introduced depending on the routing protocols. These additional delays can be taken into account by incorporating the queueing delay distribution into the cost function. The localizing node receives the signal from nodes 1 and 2 and finds the difference in the time of arrival of the signal after subtracting the known mean value of processing time at the forwarding node.

Let the positions of nodes 1, 2, and the localizing node be  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ , and  $\mathbf{P}$ ,

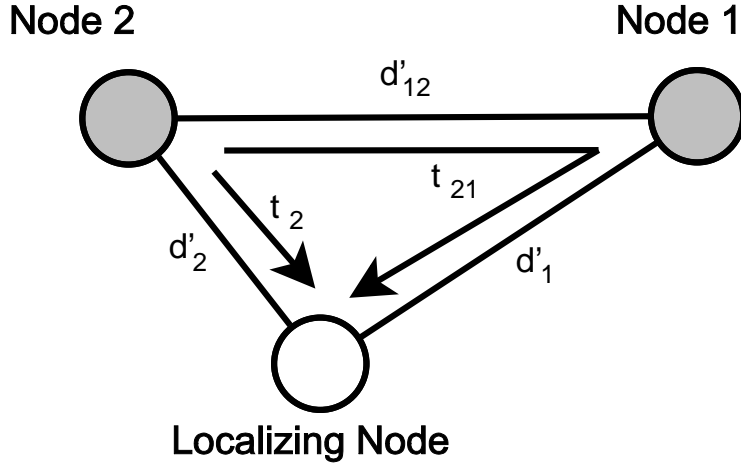


Figure 7.8: Diagram representing the basic TDoA protocol.

respectively. Denote the distances between the nodes by  $d'_{12}$ ,  $d'_1$ , and  $d'_2$ , respectively, as shown in Fig 7.8. Let the time at which the node 2 transmits its position coordinates to localizing node and node 1 be  $t_0$ . Localizing node receives the signal transmitted directly from node 2 at time  $t_2$  and the forwarded signal through node 1 at time  $t_{21}$ . Then we have:

$$\begin{aligned}
 d'_2 &= c(t_2 - t_0) \\
 d'_{12} + d'_1 &= c(t_{21} - t_0) \\
 \Rightarrow d'_1 - d'_2 &= c(t_{21} - t_2) - d'_{12} \triangleq \Delta_{21}, \tag{7.13}
 \end{aligned}$$

where  $c$  is the speed of the signal in the medium. This can be the speed of light for radio signals or the speed of sound for ultra-sonic signals. Thus, given the time difference of arrival  $t_{21} - t_2$ , the localizing node position lies on a hyperbola with foci at  $\mathbf{P}_1$  and  $\mathbf{P}_2$ .

### 7.3.1 Secure Localization Problem for TDoA

With this background on TDoA, we can now set up the secure localization problem when TDoA measurements are available. Suppose that we have  $N$  anchor nodes with known position coordinates, out of which  $L$  nodes are malicious and launch coordinated attacks. We need to determine the position of an unknown node using the time difference of arrival method to estimate the distance between the localizing node and the anchor nodes. Each pair of anchor nodes gives rise to one equation of a hyperbola. If we assume that every pair of anchor nodes is used to obtain one TDoA measurement, we have  $\binom{N}{2}$  measurements to determine the position of the localizing node, which increases as  $O(N^2)$ . In practical resource constrained networks, obtaining such a large number of measurements and solving the corresponding equations can consume a lot of resources.

To simplify the problem, we assume that there is one tamper-proof trusted anchor node in the network (say node 1) that will be used to help localize the other nodes. Each of the remaining anchor nodes transmit the beacon signal with the timestamp to the localizing node. Upon receiving the signal, the trusted node 1 forwards it to the localizing node, which thus obtains one TDoA measurement. Under this assumption, we have a manageable number of  $N - 1$  equations for the node location, which may be solved in practical resource constrained networks. Let  $\mathbf{P}_k$  denote the location of the  $k$ th anchor node. We need to determine the point  $\mathbf{P}$  that satisfies:

$$\|\mathbf{P} - \mathbf{P}_1\| - \|\mathbf{P} - \mathbf{P}_k\| = \Delta_{k1}, \quad k = 2, 3, \dots, N, \quad (7.14)$$

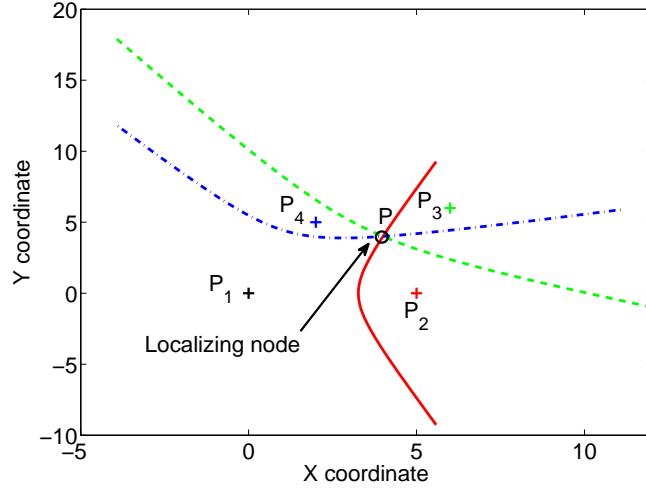


Figure 7.9: Intersection of three hyperbolas gives the location of a node when TDoA is used.

where  $\Delta_{k1}$  is the TDoA measurement obtained through anchor node  $k$  and the trusted node 1 as described in Eq. (7.13). Fig. 7.9 shows an example where four anchor nodes are located at  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ ,  $\mathbf{P}_3$ , and  $\mathbf{P}_4$ , and there are no malicious nodes or measurement noise. The hyperbolas correspond to the loci of points that are consistent with one TDoA measurement. The common intersection of the hyperbolas, denoted by point  $\mathbf{P}$ , is the position of the localizing node.

In the presence of measurement noise alone, Eq. (7.14) can be solved in the least squares sense by finding the solution to the following LS equation:

$$\begin{aligned}
 \hat{\mathbf{P}} &= \arg \min_{\mathbf{P}} \sum_{k=2}^N (\|\mathbf{P} - \mathbf{P}_1\| - \|\mathbf{P} - \mathbf{P}_k\| - \Delta_{k1})^2 \\
 &= \arg \min_{\mathbf{P}} f_{td}(\mathbf{P}).
 \end{aligned} \tag{7.15}$$

In the presence of malicious nodes, this LS solution may not be accurate. The malicious node may collude together to prevent the accurate localization of other

nodes. Based on our assumption that a tamper-proof trusted node is used to help localize the node, the attacker cannot successfully launch an attack by modifying the timestamp alone. Any changes in the timestamp corresponding to the time of transmission will not affect the distance measurement, which only depends on the difference in the time of arrival of the two signals. Instead, the strategy of attacker will be to modify the transmitted position coordinates of  $k$ th node to  $\mathbf{P}'_k$  in an intelligent way. Suppose that the attacker knows the position  $\mathbf{P}_1$  of the trusted node 1 and the position  $\mathbf{P}$  of the localizing node. Based on this knowledge, the attacker can estimate the  $t_k$  and  $t_{k1}$ -time instants at which the localizing node receives the direct beacon signal from  $k$ th node and forwarded the beacon signal from node 1. Denote by  $\mathbf{P}_{\text{mal}}$  the position where the attacker wants to shift the estimate. We then have the following relations:

$$ct_k = \|\mathbf{P}_{\text{mal}} - \mathbf{P}'_k\| \quad (7.16)$$

$$ct_{k1} = d'_{1k} + \|\mathbf{P}_{\text{mal}} - \mathbf{P}_1\| \quad (7.17)$$

where  $d'_{1k} = \|\mathbf{P}_1 - \mathbf{P}'_k\|$ . The attackers can determine a suitable value of  $P_{\text{mal}}$  and  $P'_k$  for the nodes that are compromised such that Eq. (7.15) and Eq. (7.16) are satisfied.

Our gradient descent based approach with selective pruning described in the previous section can be extended to perform secure localization in this case. The algorithm starts by randomly initializing the LS estimate  $\hat{\mathbf{P}}(0)$ . At the  $i^{\text{th}}$  step of the iteration, the gradient of the cost function  $f_{td}(\mathbf{P})$  is evaluated at the current estimate  $\hat{\mathbf{P}}(i-1)$ , and the estimate is updated by moving it one step in the direction

of the negative of the gradient, denoted by  $\mathbf{g}_{td}(i)$ :

$$\mathbf{g}_{td}(i) = - \nabla_{\mathbf{P}} f_{td}(\mathbf{P})|_{\mathbf{P}=\hat{\mathbf{P}}(i-1)},$$

The geometric interpretation of the gradient descent for TDoA is similar to the previous case for ToA, where at each iteration, a step in the direction of the negative of the gradient moves the current estimate of the location towards the intersection of the hyperbolas. At every iteration, we compute the gradient corresponding to each term in Eq. (7.15) and then sum them up to find the overall gradient. In the pruning stage, we discard a fraction of the terms with large gradient magnitude and sum up the remaining terms to obtain the gradient direction.

We perform simulations using the same settings as before. A total of  $N = 30$  nodes are distributed uniformly in a grid of size  $60m \times 60m$ . The measurement noise is assumed to be Gaussian with zero mean and  $\sigma = 2m$ . Fig. 7.10 shows the localization error under coordinated attacks by 30% of the nodes. The x-axis represents the distance  $d_a$  between the position reported by the malicious nodes and the true location. The dashed line represents the localization accuracy using the proposed method, while the solid line represents the localization accuracy using the least squares solution. From the figure, we see that the localization error using our gradient descent algorithm is less than the error obtained using the least square method under coordinated attacks, although the localization error increases with an increase in attack distance,  $d_a$ , for both approaches. The reason for this behavior can be explained by examining the probability of identifying the true location as a function of the percentage of malicious nodes. From the results shown in Fig. 7.11 for

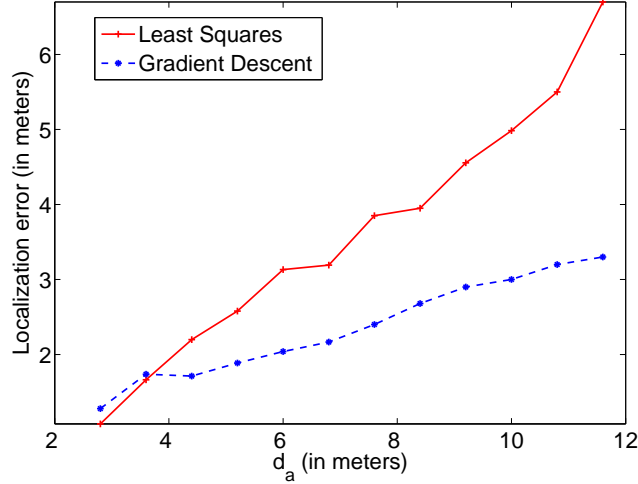


Figure 7.10: Localization accuracy for coordinated attacks by 30% of the nodes using TDoA measurements.

$d_a = 22m$ , we see that the probability of converging to the true estimate under TDoA for our scheme is approximately 10% lower than that of the ToA case of Fig. 7.6(b). When the algorithm does not converge to the correct estimate, it converges close to the position reported by the malicious nodes, which corresponds to local minimum of the cost function and incurs an error that grows with the strength of the attack. As a result, the average error increases as the distance of attack,  $d_a$ , increases. However, as compared to using the baseline LS solution, the probability of converging to the correct position is 20-30% higher for the gradient descent algorithm.

## 7.4 Chapter Summery

In this chapter, we have described a localization algorithm which takes distance estimates between a localizing entity and anchor nodes as its input. The



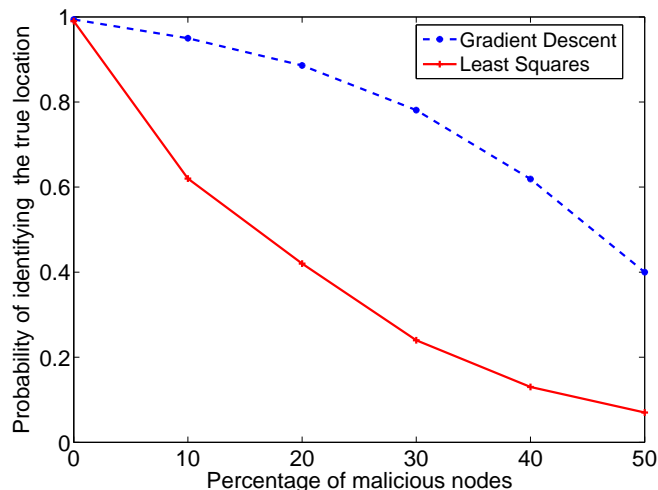


Figure 7.11: Probability of converging to the correct estimate under coordinated attacks for TDoA,  $d_a = 22\text{m}$ .

proposed algorithm is applicable to general settings in wireless sensor networks, in which the input data may be malicious due to the presence of adversaries. We proposed an iterative computationally efficient method for localization in adversarial scenarios which automatically prunes out the malicious measurements. The localization accuracy of the proposed method is better than or comparable to that of existing algorithms for secure localization in static sensor networks.

## Chapter 8

# Conclusions & Future Perspectives

In this dissertation, we have introduced a novel multimedia forensics framework and a series of techniques based on ENF signals, which can determine the time and the location of a given recording.

We have investigated mechanisms to capture ENF fluctuations in indoor lighting using optical sensors and video cameras and have demonstrated the presence of ENF signals in indoor video surveillance recordings. Recorded optical and video signals have been shown to possess a high correlation with ENF signals captured directly from power mains. The estimated time-of-recordings for sensor signals have been shown to be accurate up to a time resolution of two seconds for high signal-to-noise-ratio (SNR) optical sensors ENF signals, and that of four seconds for noisy video-ENF signals. Results from our investigations suggest that ENF signals can be used as an inherent timestamp for optical sensors and video surveillance recordings in an indoor environment. ENF signals from video clips have been shown to be robust against compression up to a bit rate of 250Kbps for standard definition

videos. ENF signals can also be used to facilitate the authentication of sensor and video data by identifying the discontinuities or alterations in ENF signals or by comparing the video-ENF signal with the reference power-ENF signal. ENF traces are also shown to provide a forensic binding of audio and visual tracks to verify their temporal synchronization and integrity.

To understand the statistical properties of ENF signals, we have proposed an autoregressive (AR) process based analytical model for these signals. We have validated the proposed model using the Box-Jenkins method and have demonstrated that the proposed model fits the ENF signals into an AR(2) model. We have used the proposed model to predict the performance of an ENF-signal based timestamp verification application for a given value of the SNR and query duration under a binary hypothesis detection framework. The trends in the receiver operating characteristics of the analytical model for different segment sizes used for matching are similar to those obtained from the experimental data. Based on the proposed model, a decorrelation based innovation process matching approach is adopted to improve the performance of the timestamp verification under the proposed framework. The experimental results with audio data demonstrated an improvement in the detection performance from 92% to 95% for a false alarm probability of 1% for 256-second long query segment, when decorrelated innovation sequences are used for matching as compared with direct matching of ENF sequences.

We have also demonstrated that the proposed AR(2) model can be used to characterize different grids. We use this property to determine the grid-region of creation of the ENF influenced recordings. We have used the AR(2) process param-

eters as features to train a classifier, which can provide 70% classification accuracy on power-ENF data and 44% accuracy for audio-ENF data on a database of recordings from seven grid regions. We have also demonstrated that the noisy nature of audio-ENF signal can cause a low classification accuracy, and it can be improved by using a noise adaptation approach based on a Bayesian multi-conditional learning framework. The audio-ENF classification improves to 64% with the proposed noise adaptation approach, as compared to 44% when no noise adaptation is performed.

We have also explored intra-grid localization capabilities of ENF signals to determine the location of a recording within the same grid-region. We have demonstrated that such location specific signatures can be extracted using a high-pass filtering mechanism from power-ENF signals. For multi-location data, we have proposed a half-plane intersection method to estimate the location of an unknown recordings. The localization accuracy from this method can be improved by adding more locations as anchor nodes. The number of constraints to estimate the feasible region increases on the order of  $\mathcal{O}(K^2)$  with the number of anchor cities  $K$ . We have also used a quantized correlation coefficient based trilateration method for localization. A combination of the half-plane intersection and the correlation quantization provides better localization accuracy as compared with using each method separately. We have also proposed a localization method for the case of general sensor networks when distance estimated between the node in question and the anchors are available. The proposed localization method is designed to be robust against adversaries injecting malicious distance information. We have demonstrated that the proposed localization approach can provide a comparable or better localization

accuracy in the presence of malicious adversaries at a much lower computational cost, as compared to the existing schemes for secure localization under malicious users.

Based on the study of this dissertation, several aspects of ENF signal analysis can be explored for multimedia forensics. As this dissertation is the first study explicating the intra-grid and reference-free grid-level localization capabilities of the ENF signal, substantial room exists for improvement in the localization capabilities of ENF signals from multimedia recordings within grid-region and at a grid-region level. To classify a recording to determine the grid-region of recording, we considered only the features obtained from the AR modeling. Other useful signal information, such as the range of deviations or sub-band decomposition, could be used for better feature representation. In addition, the grid-region location capabilities can be evaluated with a large-scale database comprising of recordings from different grid-regions. For intra-grid localization, two directions can be pursued. First, devising better ENF signal estimation techniques may further improve the SNR of the signal, which could be useful in enhancing localization capabilities. Second, different time-frequency bands of the signal can be analyzed to study their localization capabilities. This study may help in devising a better location specific feature extraction scheme.

For time-of-recording estimation tasks, the ENF signal database can be large, so it may not be feasible to correlate a given query with every possible segment of a huge database. In such cases, it is desirable to query the database by narrowing the search space using certain traits. The fundamental problem is to find a suitable representation of ENF signals by transforming them to another domain followed

by a suitable indexing mechanism. The indexing mechanism can be addressed, for example, using clustering approaches. Such an approach may make ENF querying an efficient and practical task. One related problem that needs to be addressed for large scale ENF databases is that of storage. Potential forensics and security application of the ENF analysis could require continuous recording of ENF signals for many years, meaning an efficient storage mechanism need to be devised. A study of applying lossy vs. lossless compression schemes to ENF signal storage, and the performance for applications such as timestamp estimation/verification, may provide an understanding for the tolerable amount of compression before a significant degradation in the matching performance occurs.

## Bibliography

- [1] I. J. Cox. *Digital Watermarking and Steganography*. Elsevier Science, 2008.
- [2] M.C. Stamm, M. Wu, and K.J.R. Liu. Information forensics: An overview of the first decade. *IEEE Access*, 1:167–200, 2013.
- [3] A. Swaminathan, M. Wu, and K.J.R. Liu. Component forensics. *IEEE Signal Processing Magazine*, 26(2):38–48, 2009.
- [4] C. Grigoras. Applications of ENF criterion in forensics: Audio, video, computer, and telecommunication analysis. *Forensic Science International*, 167(2-3):136–145, Apr. 2007.
- [5] R. W. Sanders. Digital authenticity using the electric network frequency. In *33rd AES International Conference on Audio Forensics, Theory and Practice*, June 2008.
- [6] M. Bollen and I. Gu. *Signal Processing of Power Quality Disturbances*. Wiley-IEEE Press, 2006.
- [7] D. P. N. Rodriguez, J. A. Apolinario, and L. W. P. Biscainho. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Transactions on Information Forensics and Security*, 5(3):534–543, Sep. 2010.
- [8] R. H. Bolt, F. S. Cooper, J. L. Flanagan, J. G. McKnight, T. G. Stockham, and M. R. Weiss. The EOB tape of June 20, 1972. In *Report on a Technical Investigation Conducted for the U.S. District Court for the District of Columbia by the Advisory Panel on White House Tapes*, May 1974.
- [9] R. Garg, A. L. Varna, and M. Wu. “Seeing” ENF: natural time stamp for digital video via optical sensing and signal processing. In *Proceedings of the 19th ACM international conference on Multimedia*, Nov. 2011.
- [10] R. Garg, A. H. Ahmad, and M. Wu. Geo-location estimation from electrical network frequency signals. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013.

- [11] A. H. Ahmad, R. Garg, and M. Wu. Instantaneous frequency estimation and localization for ENF signals. In *Proceedings of the 4<sup>th</sup> Annual Summit & Conference (APSIPA)*, Dec. 2012.
- [12] T. Harris. How light bulbs work. <http://home.howstuffworks.com/light-bulb.htm>.
- [13] A. J. C. Moreira, R. T. Valadas, and A. M. Duarte. Optical interference produced by artificial light. *Wireless Networking*, 3(2):131–140, May 1997.
- [14] W. Hernandez. Input-output transfer function analysis of a photometer circuit based on an operational amplifier. *Sensors*, 8(1):35–50, Sep. 2008.
- [15] L. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., 1993.
- [16] S. Haykin. *Advances in Spectrum Analysis and Array Processing*. Prentice-Hall, Inc., 1991.
- [17] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, Mar. 1986.
- [18] R. Roy and T. Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984–995, July 1989.
- [19] B. Friedlander. The root-MUSIC algorithm for direction finding with interpolated arrays. *IEEE Transactions on Signal Processing*, 30(1):15–29, 1993.
- [20] J. Gu, Y. Hitomi, T. Mitsunaga, and S.K. Nayar. Coded rolling shutter photography: Flexible space-time sampling. In *IEEE International Conference on Computational Photography*, Mar. 2010.
- [21] A. Bovik. *Handbook of Image and Video Processing*. Academic Press, 2000.
- [22] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck. *Discrete-time Signal Processing (2nd ed.)*. Prentice-Hall, Inc., 1999.
- [23] A. H. Ahmad, R. Garg, and M. Wu. Spectrum combining for ENF signal estimation. *IEEE Signal Processing Letters*, 20(9):885–888, 2013.
- [24] O. Ait-Aider and F. Berry. Structure and kinematics triangulation with a rolling shutter stereo rig. In *IEEE International Conference on Computer Vision*, Sep. 2009.
- [25] O. Ait-Aider, N. Andreff, M. Lavest, and P. Martinet. Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In *European Conference on Computer Vision*, May 2006.



- [26] M. Huijbregtse and Z. Geradts. Using the ENF criterion for determining the time of recording of short digital audio recordings. In *Proceedings of the 3rd International Workshop on Computational Forensics*, pages 116–124, Aug. 2009.
- [27] G. Box, G. Jenkins, and G. Reinsel. *Time series analysis: Forecasting and control*. Wiley & Sons, Inc., 2008.
- [28] R. Garg, A. L. Varna, and M. Wu. Modeling and analysis of electric network frequency signal for timestamp verification. In *IEEE International Workshop on Information Forensics and Security*, Dec. 2012.
- [29] W.-H. Chuang, R. Garg, and M. Wu. How secure are power network signature based time stamps? In *Proceedings of the 19<sup>th</sup> ACM International Conference on Computer and Communication Security*, Oct. 2012.
- [30] W.-H. Chuang, R. Garg, and M. Wu. Anti-forensics and countermeasures of electrical network frequency analysis. *Accepted for publications in the IEEE Transactions on Information Forensics & Security*.
- [31] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Inc., 2001.
- [32] R. Garg, A. L. Varna, A. H. Ahmad, and M. Wu. “seeing” enf: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing. *Information Forensics and Security, IEEE Transactions on*, 8(9):1417–1432, 2013.
- [33] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716723, 1974.
- [34] G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461464, 1978.
- [35] A. H. Ahmad, R. Garg, and M. Wu. Instantaneous frequency estimation and localization for enf signals. *To appear, 4th APSIPA Annual Summit and Conference*, Dec. 2012.
- [36] UNICEF: Child protection from violence, exploitation and abuse child trafficking. In *URL: [http://www.unicef.org/protection/57929\\_58005.html](http://www.unicef.org/protection/57929_58005.html)*.
- [37] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, 2000.
- [38] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [39] A. H. Ahmad, R. Garg, and M. Wu. ENF based location classification of sensor recordings. *Submitted to IEEE workshop on Information Forensics & Security*, 2013.

- [40] M. Ji, T.J. Hazen, J.R. Glass, and D.A. Reynolds. Robust speaker recognition in noisy conditions. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5):1711–1723, 2007.
- [41] J. Ming, P. Jancovic, P. Hanna, and D. Stewart. Modeling the mixtures of known noise and unknown unexpected noise for robust speech recognition. In *Eurospeech*, Sep. 2001.
- [42] T. F. Wu, C. J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research.*, 5:975–1005, Dec. 2004.
- [43] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in larger margin classifiers*, pages 61–74, 1999.
- [44] D. A. Reynolds. Gaussian mixture models. In *Encyclopedia of Biometric Recognition*. Springer, 2008.
- [45] T. K. Moon. The expectation-maximization algorithm. *Signal Processing Magazine, IEEE*, 13(6):47–60, 1996.
- [46] E. Parzen. On estimation of a probability density function and mode. *The Annal of Mahematical Statistics*, 33(3):1065–1076, 1962.
- [47] S. J. Tsai, L. Zhang, A. G. Phadke, Y. Liu, M. R. Ingram, S. C. Bell, I. S. Grant, D. T. Bradshaw, D. Lubkeman, and L. Tang. Frequency sensitivity and electromechanical propagation simulation study in large power systems. *IEEE Transactions on Circuits and Systems*, 54(8):1819 –1828, Aug. 2007.
- [48] D. G. Manolakis, V. K. Ingle, and S. M. Kogon. *Statistical and Adaptive Signal Processing*. McGraw-Hill, Inc., 2000.
- [49] R. Garg, A. L. Varna, and M. Wu. An efficient gradient descent approach for secure localization in resource constrained wireless sensor networks. *IEEE Transactions on Informations Forensics and Security*, 7(2):717–730, Apr. 2012.
- [50] E. Cayirci, H. Tezcan, Y. Dogan, and Y Coskun. Wireless sensor networks for underwater survelliance systems. *Ad Hoc Networks*, 4(4):431 – 446, 2006.
- [51] L. Yu, Wang. N., and X. Meng. Real-time forest fire detection with wireless sensor networks. In *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing.*, volume 2, pages 1214 – 1217, Maui, HI, USA, Sep. 2005.
- [52] E. A. Basha, S. Ravela, and D. Rus. Model-based monitoring for early warning flood detection. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys)*, pages 295 –308, Raleigh, NC, USA, 2008.

- [53] J. T. Chiang, J. J. Haas, and Y. C. Hu. Secure and precise location verification using distance bounding and simultaneous multilateration. In *Proceedings of the second ACM conference on Wireless network security*, pages 181–192, Zurich, Switzerland, 2009.
- [54] S. Capkun, K.B. Rasmussen, M. Cagalj, and M. Srivastava. Secure location verification with hidden and mobile base stations. *IEEE Transactions on Mobile Computing*, 7(4):470–483, 2008.
- [55] S. Capkun and J.-P. Hubaux. Secure positioning in wireless networks. *IEEE Journal on Selected Areas in Communications*, 24(2):221–232, Feb 2006.
- [56] D. Liu, P. Ning, A. Liu, C. Wang, and W. K. Du. Attack-resistant location estimation in wireless sensor networks. *ACM Transactions on Information and System Security*, 11(4):1–39, 2008.
- [57] L. Lazos and R. Poovendran. SeRLoc: secure range-independent localization for wireless sensor networks. In *Proceedings of the 3rd ACM Workshop on Wireless Security (WiSe)*, pages 21–30, Philadelphia, PA, USA, 2004.
- [58] R. O. Duda and P. E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.
- [59] Z. Li, W. Trappe, Y. Zhang, and B. Nath. Robust statistical methods for securing wireless localization in sensor networks. In *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN)*, page 12, Los Angeles, CA, USA, 2005.
- [60] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [61] Y. Takizawa, P. Davis, M. Kawai, H. Iwai, A. Yamaguchi, and S. Obana. Self-organizing location estimation method using received signal strength. *IEICE Transactions on Communications*, B89-B(10):2687–2695, 2006.
- [62] P. Rousseeuw and K. V. Driessen. Computing lts regression for large data sets. *Data Mining and Knowledge Discovery*, 12(1):29–45, 2006.
- [63] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, Inc., 1987.
- [64] R. Garg, A. L. Varna, and M. Wu. Gradient descent approach for secure localization in resource constrained wireless sensor networks. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1854–1857, Dallas, TX, USA, Mar. 2010.

- [65] S. Yi, R. Wheeler, Y. Zhang, and M. Fromherz. Localization from mere connectivity. *Proceedings of ACM International Symposium on Mobile Ad-hoc Networking & Computing*, pages 201–212, 2003.
- [66] R. H. Kwong and E. W. Johnston. A variable step size LMS algorithm. *IEEE Transactions on Signal Processing*, 40(7):1633 –1642, Jul. 1992.
- [67] S. Zhong, M. Jadliwala, S. Upadhyaya, and C. Qiao. Towards a theory of robust localization against malicious beacon nodes. In *Proceedings of the 27th IEEE International Conference on Computer Communication (INFOCOM)*, Los Angeles, CA, USA, 2008.
- [68] N. Patwari. Location estimation in sensor networks. *PhD Thesis, University of Michigan*, 2005.
- [69] D. Munoz, F. Bouchereau, C. Vargas, and R. Enriquez. *Position Location Techniques and Applications*. Elsevier, 2009.