



**IJCER**

ONLINE PEER REVIEWED JOURNAL

International Journal of Computational  
Engineering research

# **International Journal of Computational Engineering Research**

## **Volume, 4, Issue, 3 March, 2014**

**Published by: IJCER**

**ISSN: 2250–3005 (online version)**

**[www.ijceronline.com](http://www.ijceronline.com)**

# Editorial Board

## Editor-In-Chief

### **Prof. Chetan Sharma**

Specialization: Electronics Engineering, India  
Qualification: Ph.d, Nanotechnology, IIT Delhi, India

## Editorial Committees

### **DR.Qais Faryadi**

Qualification: PhD Computer Science  
Affiliation: USIM(Islamic Science University of Malaysia)

### **Dr. Lingyan Cao**

Qualification: Ph.D. Applied Mathematics in Finance  
Affiliation: University of Maryland College Park,MD, US

### **Dr. A.V.L.N.S.H. HARIHARAN**

Qualification: Phd Chemistry  
Affiliation: GITAM UNIVERSITY, VISAKHAPATNAM, India

### **DR. MD. MUSTAFIZUR RAHMAN**

Qualification: Phd Mechanical and Materials Engineering  
Affiliation: University Kebangsaan Malaysia (UKM)

### **Dr. S. Morteza Bayareh**

Qualificatio: Phd Mechanical Engineering, IUT  
Affiliation: Islamic Azad University, Lamerd Branch  
Daneshjoo Square, Lamerd, Fars, Iran

### **Dr. Zahéra Mekkioui**

Qualification: Phd Electronics  
Affiliation: University of Tlemcen, Algeria

### **Dr. Yilun Shang**

Qualification: Postdoctoral Fellow Computer Science  
Affiliation: University of Texas at San Antonio, TX 78249

### **Lugen M.Zake Sheet**

Qualification: Phd, Department of Mathematics  
Affiliation: University of Mosul, Iraq

### **Mohamed Abdellatif**

Qualification: PhD Intelligence Technology  
Affiliation: Graduate School of Natural Science and Technology

**Meisam Mahdavi**

Qualification: Phd Electrical and Computer Engineering

Affiliation: University of Tehran, North Kargar st. (across the ninth lane), Tehran, Iran

**Dr. Ahmed Nabih Zaki Rashed**

Qualification: Ph. D Electronic Engineering

Affiliation: Menoufia University, Egypt

**Dr. José M. Merigó Lindahl**

Qualification: Phd Business Administration

Affiliation: Department of Business Administration, University of Barcelona, Spain

**Dr. Mohamed Shokry Nayle**

Qualification: Phd, Engineering

Affiliation: faculty of engineering Tanta University Egypt

## CONTENTS:

S.No.	Title Name	Page No.
<b>Version I</b>		
1.	Study of Combustion Characteristics of Fuel Briquettes <b>Amit Talukdar, Dipankar Das, Madhurjya Saikia</b>	01-03
2.	The Birefringent Property of an Optical Resin for the Study of a Stress Field Developed in a Three Point Loading Beam <b>K. Touahir , A. Bilek S. Larbi , S. Djebali</b>	04-08
3.	Optimization of Image Search from Photo Sharing Websites <b>Shubhada Mali, Sushama Kadam, Ganesh Shelke, Rajani Ghodake</b>	09-12
4.	Manets: Increasing N-Messages Delivery Probability Using Two-Hop Relay with Erasure Coding <b>A.Vijayalakshmi, J.Praveena</b>	13-19
5.	<b>Survey paper on Virtualized cloud based IPTV System</b> Jyoti P. Hase, Prof. R.L.Paikrao	20-24
6.	<b>The Search of New Issues in the Detection of Near-duplicated Documents</b> Hasan Naderi , Narges salehpour, Mohammad Nazari farokhi, Behzad Hosseini chegeni	25-31
7.	<b>Mathematical Modeling of Class B Amplifier Using Natural and Regular Sampled Pwm Modulation</b> N. V. Shiwarkar, K. G. Rewatkar	32-37
8.	<b>Neighborhood Triple Connected Domination Number of a Graph</b> G. Mahadevan, N. Ramesh, C. Sivagnanam, Selvam Avadayappan, A. Ahila, T.Subramanian	38-48

## Version II

1.	Bioelectrical Impedance Analysis (BIA) For Assessing Tbw and Ffm of Indian Females <b>Munna Khan, Shabana Mehfuz, Ghazala PerveenKhan</b>	01-17
2.	The Applications of Computational Intelligence (Ci) Techniques in System Identification and Digital Filter Design <b>Jainarayan Yadav, Sanjay Gurjar</b>	18-24
3.	Bi-Level Weighted Histogram Equalization with Adaptive Gamma Correction <b>Jeena Baby, V. Karunakaran</b>	25-30
4.	Ultraspherical Solutions for Neutral Functional Differential Equations <b>A. B. Shamardan , M. H. Farag, H. H. Saleh</b>	31-42
5.	Solid Waste Management in Mahaboobnagar Municipality <b>Dr. C.Sarala , G.SreeLakshmi</b>	43-52
6.	Survey Paper on Creation Of dynamic query Form for mining highly optimized transactional databases <b>Jayashri M. Jambukar</b>	53-57
7.	Image Performance Tuning Using Contrast Enhancement Method with Quality Evaluation <b>Alphy George, S John Livingston</b>	58-63
8.	Customized Query Interface Integration using Attribute Constraint Matrices <b>Sherlin Mary Koshy, Belfin R. V.</b>	64-67
9.	Analysis of Image Segmentation Methods Based on Performance Evaluation Parameters <b>Monika Xess , S. Akila Agnes</b>	68-75

10.	Controlling Computer Operations using Brain-Wave Computing <b>Shanmugapriya.B , Akshaya.T , Kalaivani.K , Anbarasu.V</b>	76-81
11.	A Novel Approach for the Effective Detection of Duplicates in XML Data <b>Anju Ann Abraham, S. Deepa Kanmani</b>	82-87
12.	Survey Paper on Integrity Auditing of Storage <b>Ugale Santosh A</b>	88-92
13.	Lookup table embedded in (FPGA) for network security <b>ZuhirNemerAlaaraj , Abdelrasoul Jabar Alzubaidi</b>	93-96
14.	Survey of Steganalysis Technique for Detection of Hidden Messages <b>Vanita J. Dighe,Prof. Baisa L. Gunjal</b>	97-100

# Study of Combustion Characteristics of Fuel Briquettes

Amit Talukdar<sup>1</sup>, Dipankar Das<sup>2</sup>, Madhurjya Saikia<sup>3</sup>

<sup>1</sup> Asst Prof., Mechanical Engineering Department, Dibrugarh University Institute of Engineering and Technology, Dibrugarh, Assam,

<sup>2</sup> Asst Prof., Mechanical Engineering Department, Dibrugarh University Institute of Engineering and Technology, Dibrugarh, Assam,

<sup>3</sup> Asst Prof., Mechanical Engineering Department, Dibrugarh University Institute of Engineering and Technology, Dibrugarh, Assam,

## ABSTRACT.

Biomass material such as rice straw, banana leaves and teak leaves (*Tectona grandis*) are densified by means of wet briquetting process at lower pressures of 200-1000 kPa using a piston press. Wet briquetting is a process of decomposing biomass material up to a desired level under controlled environment in order to pressurize to wet briquettes or fuel briquettes. Upon drying these wet briquettes could be used as solid fuels. This study is aimed at to determine combustion characteristics of briquettes which will facilitate to answer some of the questions regarding the usefulness of fuel in terms of production of harmful gases and fly ashes during combustion which are common indoor air pollutants in many households and effectiveness of the fuel in terms of heat value.

**KEYWORDS:** Biomass, fuel briquettes, calorific value, combustion rate.

## I. INTRODUCTION

With rapid depletion of world petroleum reserves and increased demand of petroleum products, especially of transportation fuels, it has posed a serious problem of energy crisis in India. Rural India is mostly affected by it. Therefore, there is a great demand of cheap and easily available fuel for cooking in rural household sector. In this regard, briquetting could be a viable option. Biomass briquetting is the densification of loose biomass material to produce compact solid composites of different sizes with the application of pressure. Three different types of densification technologies are currently in use. The first, called pyrolyzing technology relies on partial pyrolysis of biomass, which is added with binder and then made into briquettes by casting and the first, called pyrolyzing technology relies on partial pyrolysis of biomass, which is added with binder and then made into briquettes by casting and pressing. The second technology is direct extrusion type, where the biomass is dried and directly compacted with high heat and pressure. The last type is called wet briquetting in which decomposition is used in order to breakdown the fibers. On pressing and drying, briquettes are ready for direct burning or gasification. These briquettes are also known as fuel briquettes. Fuel briquettes are easy to manufacture at a very cheap cost. This technique has gained popularity in some of the African and Asian countries [1].

## II. METHOD AND MATERIAL

India produces large amounts of bio waste material every year. This includes rice straw, wheat straw, coconut shells and fibres, rice husks, stalks of legumes and sawdust. Some of this biomass is just burnt in air; some like rice husk are mostly dumped into huge mountains of waste. Open-field burning has been used traditionally to dispose of crop residues and sanitize agricultural fields against pests and diseases. Based on wide availability in country side and composition such as lesser amount of lignin and ash content compared to other available biomass, rice straw, banana leaves and teak leaves are selected for wet briquetting [2]. The first step of wet briquetting process is decomposition of biomass material up to a desired level in order to pressurize to wet briquettes at a lower pressure [3, 4]. The biomass materials are chopped in sizes about 10 mm. The chopped biomass materials are soaked in water and kept in open at an ambient temperature. At regular interval of days, hand test such as shake test is performed to check the desired level of decomposition in biomass materials. The good briquette material does not fall apart when held over the upper 1/2 portion and shaken vertically a few times in the hand test [3]. Decomposition loosens fibers of biomass materials. Time requirements for desired level of decomposition vary with biomass types. Rice straw and teak leaves take 19 days while banana leaves take 28 days to reach the desired status. The wet biomass which reaches optimum level of decomposition are pressurized by manually operated piston press of internal diameter 45 mm at varying pressure level ranging from 200 kPa to 1000 kPa. A dwell time of 40 seconds is observed during briquette formation [5]. Even after pressurization, the briquettes are wet and therefore they need to be dried up to 8% moisture content.

In the second step, we blended wet biomass of rice straw, banana leaves and teak leaves with *Mesua ferrea* L. seed cakes which are a waste material after extraction of vegetable oil for biodiesel production from *Mesua ferrea* L. seeds. Proximate analysis, calorific values combustion rates of briquettes are done to assess effectiveness of briquettes as cooking fuel. For proximate analysis and calorific value determination, ASTM E870 - 82(2013) and ASTM D2015-00 methods are used respectively. But in order to assess burning rate at room condition a simple test as proposed by Joel Chaney [6] is applied. To determine Combustion rate or burning rate which is mass loss per unit time, the briquettes are dried at 105°C so that it does not affect on combustion or burning. Dried briquette is placed on a steel wire mesh grid resting on three supports allowing free flow of air. Now the whole system is placed on mass balance. Briquette is ignited from top and mass loss data is taken at an interval of 30 seconds.

### III. RESULTS AND DISCUSSIONS

#### 3.1 Proximate analysis

Proximate analysis of briquettes samples is done and results are presented in Table 1 and 2. From the analysis, it is found that volatile matter in Nahar added briquettes is more corresponding to non added ones. It is due to oil content of nahar (*Mesua ferrea* L.) seed cake as the seed cake contains 14 % even after oil expulsion.

Table 1: Proximate analysis of rice straw, banana and teak leaves briquettes

Proximate analysis	Rice straw Briquettes (% dry fuel)	Banana leaves briquettes (% dry fuel)	Teak leaves briquettes (% dry fuel)
Fixed carbon	15.59-19.86	12.16-12.63	8.88-9.21
Volatile matter	66.45-68.59	79.2-80.05	77.32-77.88
Ash	12.74-18.68	10.1-13.4	17.27-18.08

Table 2

Proximate analysis of nahar seed cake added briquette samples

Proximate analysis	Rice straw briquettes (% dry fuel)	Banana leaves briquettes (% dry fuel)	Teak leaves briquettes (% dry fuel)
Fixed carbon	7.79-9.88	8.02-10.52	10.49-13.1
Volatile matter	82.32-90.5	83.39-87.625	75.75-86.45
Ash	7.65-9.88	7.51-9.10	8.9-13.55

#### 3.2 Calorific value

Calorific values of rice straw, banana leaves and teak leaves are determined to be 13.57 MJ/kg, 14.98 MJ/kg, and 11.78 MJ/kg respectively. However, when nahar seed cake is added with the biomass materials, calorific values are found to increase in each case. Nahar added rice straw, banana leaves and teak leaves briquettes have calorific values of 19.76 MJ/kg, 19.21 MJ/kg and 19.10 MJ/kg respectively. This is due combine effect of mixing. Calorific value of *Mesua ferrea* L. seed cake is 19.6 MJ/kg. The percentage increase in calorific values in nahar seed cake added rice straw briquette samples are found more than that of the other two nahar seed cake added samples.

#### 3.3 Combustion rate

Fig.1 shows the variation of combustion rates of briquettes over applied die pressures of briquette production. With increase in applied die pressure, briquettes become more compact and hence combustion rate is reduced. It may be due to increase in density, voids which contain trapped air decrease. It is observed that nahar added teak leaves have the highest combustion rate.



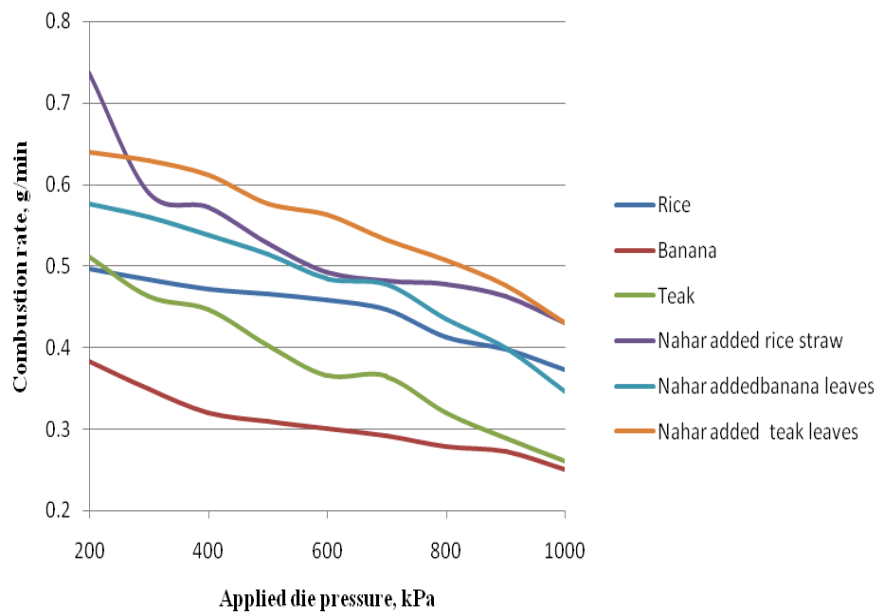


Fig.1 Variation combustion rate of briquettes over applied die pressure

#### IV. CONCLUSION

With the help of this study we aim to standardize the practice of briquette making for the commercial purpose from the combustion characteristics point of view. In the present study, the burning rates of briquettes in room condition are assessed to arrive at an conclusion increasing applied die pressure negatively effect combustion characteristic like burning rate as density of briquettes increases which in turn reduces porosity of the briquettes . It is also noticed that combustion characteristics are improved on blending of nahar seed cakes with the existing biomasses. It has raised calorific value and burning rate of briquettes. Nahar added rice straw shows the highest combustion rate of 0.43 g/min. Apart from that proximate analysis and calorific value tests will so help us to give answer to some of the questions regarding the fuel such as whether it produces too much harmful fly ash and unwanted gases which are general indoor air pollutants many households and effectiveness of the fuel in terms of heat value. A fuel without heat value would be useless as a lot of fuel will be needed during use for its lower heat value.

#### REFERENCES

- [1] Saikia M. and Baruah D.C. Analysis of Physical Properties of Biomass Briquettes Prepared by wet briquetting method, IJERD, vol 6, issue 5, pg 12-14.
- [2] Johnson Sarah, Brar D.S. and Buresh R.J. . Faster anaerobic decomposition of a brittle straw rice mutant: Implications for residue management, Soil Biology & Biochemistry, vol 38, pg 1880–1892(2006).
- [3] Liang C and McClendon R. W. The influence of temperature and moisture contents regimes on the aerobic microbial activity of a bio solids composting blend, Bio resource Technology, vol 86, pg 131–137 (2003).
- [4] Stanley R. Fuel Briquette making. Legacy Foundation (2003).
- [5] Chin O.C., Siddiqui K.M. Characteristics of some biomass briquettes prepared under modest die pressures. Biomass and Bioenergy 2000; 18: 223-228.
- [6] Chaney J. O., Clifford M. J., Wilson R. An experimental study of the combustion characteristics of low-density biomass briquettes. Biomass magazine 2010, Vol 1.

# The Birefringent Property of an Optical Resin for the Study of a Stress Field Developed in a Three Point Loading Beam

<sup>1</sup>K. Touahir\*, <sup>1</sup>A. Bilek, <sup>2</sup>S. Larbi and <sup>2</sup>S. Djebali

<sup>1</sup>Département de Génie Mécanique, Université Mouloud Mammeri, Tizi-Ouzou, Algérie

<sup>2</sup>L.E.M.M., Université Mouloud Mammeri, Tizi-Ouzou, Algérie

## ABSTRACT

The birefringent property, known also as the double-refraction phenomenon, is used in a polariscope to study a stress field developed in a three point loading beam. The model used for this analysis was made of an epoxy resin (PLM4) and a hardener (PLMH). The stress field was locked in the model by the stress freezing technique. Photoelastic fringes obtained on the analyzer of a regular polariscope were used to determine completely the stress field. A finite elements analysis was also conducted in order to determine the stress field numerically. A whole field comparison of the experimental photoelastic fringes and the simulated ones and a local analysis using the principal stresses difference showed very good agreement between the experimental solution and the numerical one.

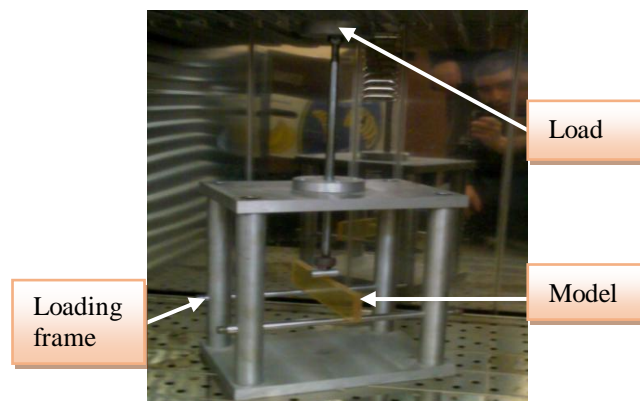
**KEYWORDS:** Birefringence, photoelasticity, fringe, stress.

## I. INTRODUCTION

The analysis of stress fields is of very high importance in the design of machinery components. Various methods can be used to solve this kind of problem [1-8], analytical as well as experimental. However, analytical solutions may, in some cases, be difficult to develop. Here, a numerical solution using the finite elements analysis was used to obtain the solution. Stresses can be obtained rapidly in any desired position of the model. To validate the finite elements solution tests were conducted on a regular polariscope (fig.1) which uses the birefringence phenomenon to analyze stress fields. The isochromatic and isoclinic fringe patterns are used to determine stresses.

## II. EXPERIMENTAL ANALYSIS

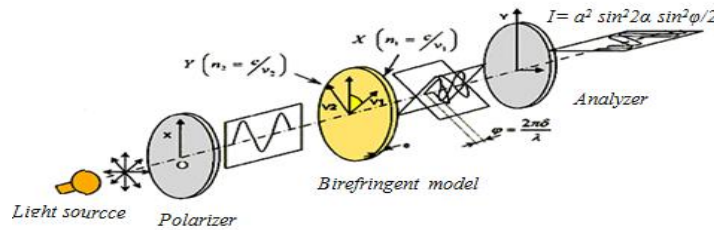
A three point loading beam is analyzed in a plan polariscope which allows observing two types of photoelastic fringes, the isochromatics and the isoclinics. The isochromatics are loci of points having the same shear stress and the isoclinics, which appear in dark color, are loci of points for which the principal stresses directions are parallel to the polarizer and the analyzer axes. The stress field can therefore be completely determined, principal stresses with their directions can be obtained in the whole model. The three point loading beam model (Figure 1) is mounted on a loading frame inside an oven at the stress freezing temperature of 130°C in order to lock the stress field inside the model. The model will then be analyzed on a regular polariscope with plan polarized light and with circularly polarized light in order to obtain the isochromatic fringe pattern and the isoclinic fringe pattern.



**Figure 1:** The model mounted on the loading frame inside an oven at the stress freezing temperature

To help the reader, a brief review of the experimental method is given below. Figure 2 shows the well-known photoelastic method based on the birefringent phenomenon. The light intensity obtained on the analyzer after traveling through the polarizer, the model and the analyzer is given by the following relation (1). The terms  $\sin^2 2\alpha$  and  $\sin^2 \phi/2$  give respectively the isoclinic fringes and the isochromatic fringes.

$$I = a^2 \sin^2 2\alpha \sin^2 \phi/2 \tag{1}$$



**Figure 2:** Light propagation through a photoelastic model

The light wave length used is  $\lambda=546\text{nm}$ . The isochromatic fringes allowed us to obtain the values of the principal stresses difference on the model by using the well known relation (2) where  $\lambda$  is the light wave length and  $C$  the optical characteristic of the birefringent material:

$$\sigma_1 - \sigma_2 = N (\lambda/C)/e \tag{2}$$

This can only be done if the values of the fringe orders have been completely determined. The values of the fringe order  $N$  are determined either by the compensation technique or, whenever possible, by starting from a non stressed region on the model for which  $N=0$ . The fringe orders can then be easily deduced for the other fringes. The ratio  $f=\lambda/C$  called the fringe constant depends on the light wave used and the model material. Several solutions are available to obtain this value easily. Here, we subjected a disc (diameter  $D = 48 \text{ mm}$  and thickness  $e = 14 \text{ mm}$ ) to a compressive load ( $P= 15\text{N}$ ). The value of the principal stresses difference is given by the well known relation (3):

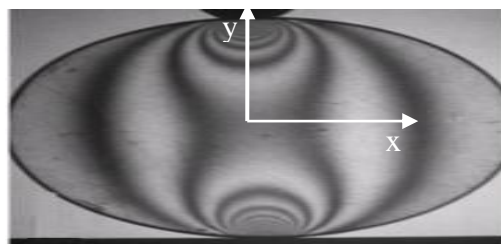
$$\sigma_1 - \sigma_2 = \frac{8P}{\pi e D} \frac{(D^4 - 4D^2 x^2)}{(D^2 + 4x^2)^2} = Nf / e \tag{3}$$

The value of the fringe constant can therefore be easily determined with the following relation (4) for different positions along the horizontal axis:

$$f = \frac{8P}{\pi D N} \frac{(D^4 - 4D^2 x^2)}{(D^2 + 4x^2)^2} \tag{4}$$

Where  $x$  represents the position of the isochromatic fringe. This distance  $x$  is taken along the diameter of the disc starting from the center of the disc (figure 3).

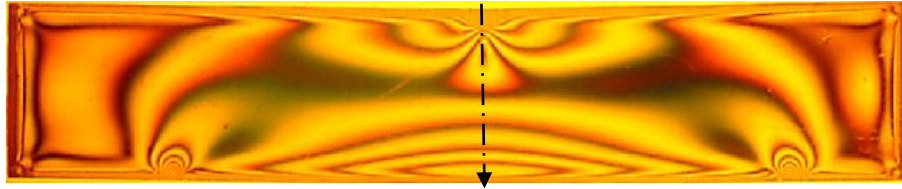
The obtained value ( $f=0.36 \text{ N/mm/fringe}$ ) was then used in the experimental solution as well as in the numerical solution to determine the stress values in the neighborhood of the contact zone along the direction of the applied load.



**Figure 3:** Isochromatics fringes obtained on the disc

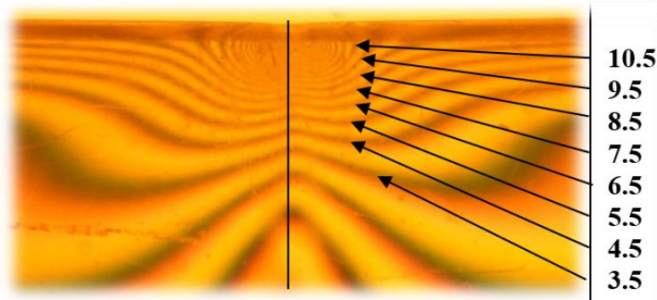
### 2.1 Experimental results

The isochromatic fringes obtained with circularly polarized light (Figure 4), can be used for comparison purposes with the finite element solution and also to determine the experimental values of the principal stress difference in the neighborhood of the contact zones, along the vertical axis, for comparison purposes with the numerical values obtained with the finite element solution. We can see, as expected, a concentration of stresses close to the contact zone.



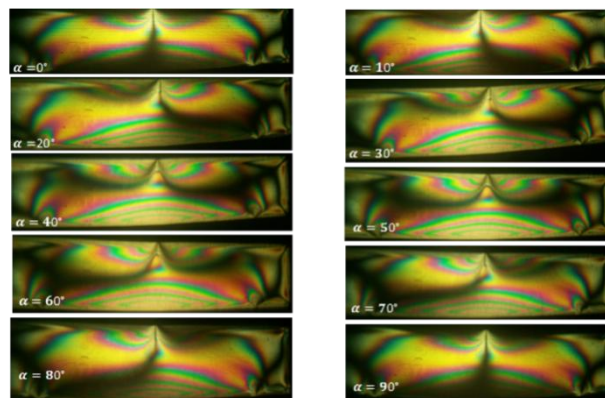
**Figure 4:** Experimental isochromatic fringes

A zoom (figure 5) allows to see clearly the different fringes in the neighborhood of the contact zone. We can therefore obtain the different values of the fringe orders which are necessary to obtain the graph of the principal stress difference along the vertical axis.



**Figure 5:** Fringe order values in the neighborhood of the contact zone

The isoclinics, which are the dark fringes, were obtained with plane polarized light for different polarizer and analyzer positions, the quarter wave plates were removed from the light path. We can see clearly the dark isoclinic fringe as it moves on the model. These isoclinics can be used to obtain the isostatics which are the principal stresses directions. The isostatics give a good understanding of how stresses are distributed in the volume of the model. This is very a important result for the design of mechanical components.



**Figure 6:** Isoclinics recorded for different polarizer and analyzer positions

### III. NUMERICAL ANALYSIS

For a first approach of the solution we consider that the material behaves everywhere as a purely elastic isotropic material. Fringe constant  $f=0.36$  N/mm/fringe, Young's modulus ( $E=15.9$ MPa) and Poisson's ratio ( $\mu_2=0.4$ ) were introduced in the finite element program. The mesh was refined in the neighborhood of the contact zone (figure 7) in order to achieve better approximation of stresses.

**3.1. Numerical calculation of the isochromatic and the isoclinic fringes**

The following relation (5) which can be obtained readily from Mohr's circle for stresses allows us to calculate the principal stresses difference at any point of a stressed model.

$$((\sigma_x - \sigma_y)^2 + 4\tau_{xy}^2)^{0.5} = \sigma_1 - \sigma_2 = Nf/e \tag{5}$$

The different values of the retardation angle  $\varphi$  can be calculated at any point on the model using the following relation (6):

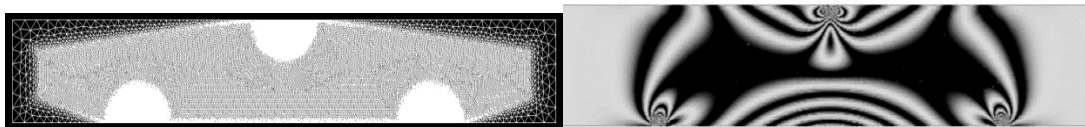
$$\varphi = 2\pi N = 2\pi e/f ((\sigma_x - \sigma_y)^2 + 4\tau_{xy}^2)^{0.5} \tag{6}$$

The different values of  $\sin^2\varphi/2$  which represents the simulated isochromatic fringes (figure 7) have been easily calculated. A comparison can then be made with the isochromatic fringes obtained experimentally with a monochromatic light (figure 4). We can see relatively good agreement; however in the neighborhood of the contact zone we can see some discrepancies. Another comparison using the principal stresses difference (figure 9) shows relatively good agreement even though close to the contact zone it was difficult to obtain experimentally the stresses.

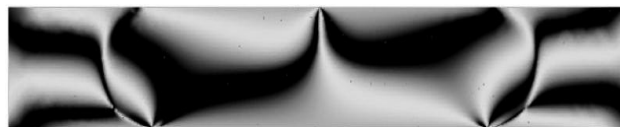
The term  $\sin^2 2\alpha$  represents the isoclinic fringes which are loci of points where the principal stresses directions are parallel to the polarizer and the analyzer. In the simulation program the different values of the isoclinic parameter  $\alpha$  can be calculated with the following relation (eq. 7) which can be obtained readily from Mohr's circle of stresses:

$$\alpha = \arctan(2\tau_{xy} / (\sigma_x - \sigma_y)) \tag{7}$$

The different values of  $\sin^2 2\alpha$  can therefore be calculated and displayed (figure 8). The comparison is then possible with the experimental isoclinic fringes which are the dark fringes obtained experimentally on figure 6. Relatively good agreement can be observed between the experimental and the simulated isoclinic dark fringes. The simulated fringes were obtained only for  $\alpha=60^\circ$ .

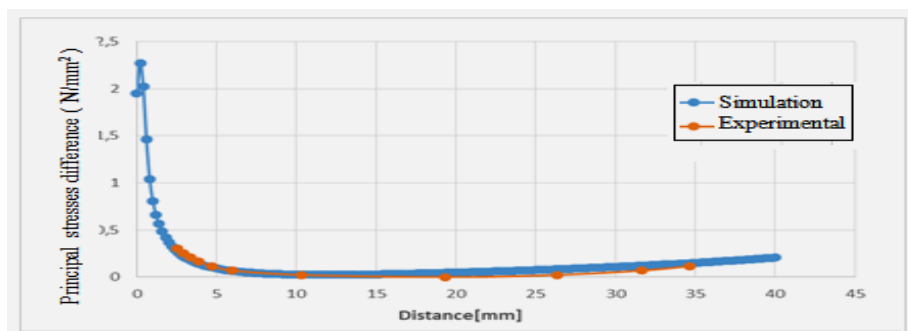


**Figure 7:** Finite element meshing and calculated isochromatic fringes



**Figure 8:** Calculated isoclinic for  $\alpha=60^\circ$

Relatively good agreement are observed between the experimental fringes (figure 4) and the simulated fringes (figure 7 right).



**Figure 9:** Principal stresses difference along the vertical axis osymmetry

#### IV. CONCLUSION

We have used the birefringent phenomenon to analyze stresses on a three point loading beam. The photoelastic fringes obtained on the analyzer were used to determine completely the stress field. A finite element solution was developed in order to simulate the photoelastic fringes and the stress values on the loaded model. Relatively good agreements were obtained between the experimental solution and the finite elements simulation.

#### REFERENCES

- [1] Germaneau, A., Peyruseigt, F., Mistou, S., Doumalin, P. Dupré, J.C., Experimental study of stress repartition in aeronautical spherical plain bearing by 3D photoelasticity : validation of a numerical model, 5th BSSM International Conference on Advances in Experimental Mechanics, Septembre 2007, University of Manchester, UK.
- [2] A. Mihailidis, V. Bakolas, & N. Drivakovs, Subsurface stress field of a dry line Contact. *Wear* V. 249, I.7, pp 546 –556, 2001.
- [3] A. Bilek, J.C. Dupré, A. Ouibrahim, & F. Bremand, 3D Photoelasticity and numerical analysis of a cylinder/half-space contact problem, *Computer Methods and Experimental Measurements for Surface Effects and Contact Mechanics*, VII, pp 173 -182, 2000.
- [4] A. Mihailidis, V. Bakolas and N. Drivakovs. Subsurface Stress Field of a Dry Line Contact, *Wear* Volume 249, Issue 7, July 2001 pp. 546 –556.
- [5] Budimir Mijovic and Mustapha Dzoclo. Numerical Contact of a Hertz Contact Between two Elastic Solids, *Engineering Modeling* 13 (2000) 3-4, pp. 111-117.
- [6] A. Bilek, F. Djeddi, Photoelastic and numerical stress analysis of a 2D contact problem and 3D numerical solution for the case of a rigid body on a deformable one. *WIT Transaction on Modeling and Simulation*, Vol 51, © 2011 pp 177-187, WIT Press
- [7] J. W Dally and F. W. Riley, *Experimental stress analysis*. McGraw-Hill, Inc, 1991.
- [8] A. Bilek, J.C. Dupré , A. Ouibrahim, F. Bremand, Non-destructive automated photoelastic analysis of a stress field with unknown prescribed boundary conditions, 3rd International Conference, Advanced Composite Materials Engineering COMAT 2010, 27- 29 October 2010, Brasov, Romania, ISSN 1844-9336.
- [9] McKelvie, J. (1998), Moiré strain analysis: an introduction, review and critique, including related techniques and future potential. *J. Strain Analysis*, 33 (2) 137-151.
- [10] R. Rotinat, V. Valle, and Cottron C (1998). Nouvelle technique de création de réseau pour la mesure optique de grandes déformations. *Proc. Photomécanique 98 – Etude du comportement des matériaux et des structures*, Marne.
- [11] Dupré J.C. (1992) Traitement et analyse d'images pour la mesure de grandeurs cinématiques, déplacement et déformations à partir de la granularité laser et de réseaux croisés, et pour l'étude de couplages thermomécaniques.
- [12] Brémand F., Dupré J.C., and Lagarde A. (1995) Mesure des déformations sans contact par analyse d'images. *Proc. Photomécanique 95 - Etude du comportement des matériaux et des structures*, Cachan, 171-177.

## Optimization of Image Search from Photo Sharing Websites

Shubhada Mali<sup>1</sup>, Sushama Kadam<sup>2</sup>, Ganesh Shelke<sup>3</sup>, Rajani Ghodake<sup>4</sup>

<sup>1, 2, 3, 4</sup> S.V.P.M's COE Malegaon (Bk). Department Of Computer Engg.

### ABSTRACT

The social networking sites, such as flicker allows users to upload images and annotate it with descriptive labels known as tags. Personalized image searching is the way to searching images according to intension of users and that personalized image result is relevant to the individual user. Personalized web search takes an advantage of information about an individual that tagging to an image for identifying the most relevant image result for that person. The main challenge for personalization lies in collecting user profiles which describes information about the user. The user preferences and fired query are used to obtained relevant image result. The proposed system contains three components: A Ranking based multi-correlation tensor factorization (RMTF) model is proposed to perform annotation prediction, which is considered as user's preference according to annotating or tagging to an image. Corpus is used to analyze users, their annotating images and users tags for each image to find users specific topics .The proposed algorithm perform topic modeling which is used to generate user specific topics. The single word query selection is used for searching relevant image result. The query mapping or query relevance and topic sensitive user preferences (TSUP) are integrated into final ranked result of relevant images.

**KEYWORDS:** Relevant search, RMTF, image annotation, user preferences, user specific topics, query relevance. TSUP.

### I. INTRODUCTION

Keyword-based search has been the most popular search in today's searching world. The result of Keyword based search is better than Google .On Google search engine user or searcher did not find relevant image result. This is because of two reasons.

1) Queries are in general short and non-specific.

2) Different users may have different intentions for the same query

Searching for apple by a farmer has a different meaning from searching by a technical person .There is one solution to solve these problems is personalized search where user specific information is considered to distinguish between exact intentions of user queries and re-ranked the images.



Fig. 1: (top) non-personalized and (bottom) personalized search results for the query "Samsung Laptop".

Fig. 1 shows the example for non-personalized and personalized image search results from the search engines. The non-personalized search returned results only based on the user query relevance and displays Samsung laptop images as well as it can displays the Samsung charger battery on the above image in fig.1. While personalized search results consider as both user query relevance and user preference, so the personalized results from a laptop lover rank the laptop images on the top. Increasingly developed social networking websites, like Flickr and YouTube allow users to create, share, upload, and annotate images. Flickr database is used to demonstrate the effectiveness of proposed system. The proposed system has two components

- 1) Ranking Based Multi-correlation Tensor Factorization model (RMTF) is used to calculate user's annotation prediction which provides user preferences to assigning tag on image. RMTF avoids common noisy problem and sever sparsity problem.
- 2) User Specific Topic Modeling (USTM) is introduced for performing topic modeling .Mapping query relevance and user preferences are combined into providing highly relevant ranked images.

## II. BACKGROUND

Existing system	Merits	Demerits
Personalizing Search via Automated Analysis of Interests and Activities	Personalized images obtained.	User's profile is not created. Ranked result is not displayed.
Discovering and using groups to improve personalized search	Personalized images obtained.	User's profile is not created. Ranked result is not displayed.
Personalized search by tag based user profile and resource profile in collaborative tagging systems	Personalized images obtained. User's profile is created.	Ranked result is not displayed.

## III. PROPOSED SYSTEM

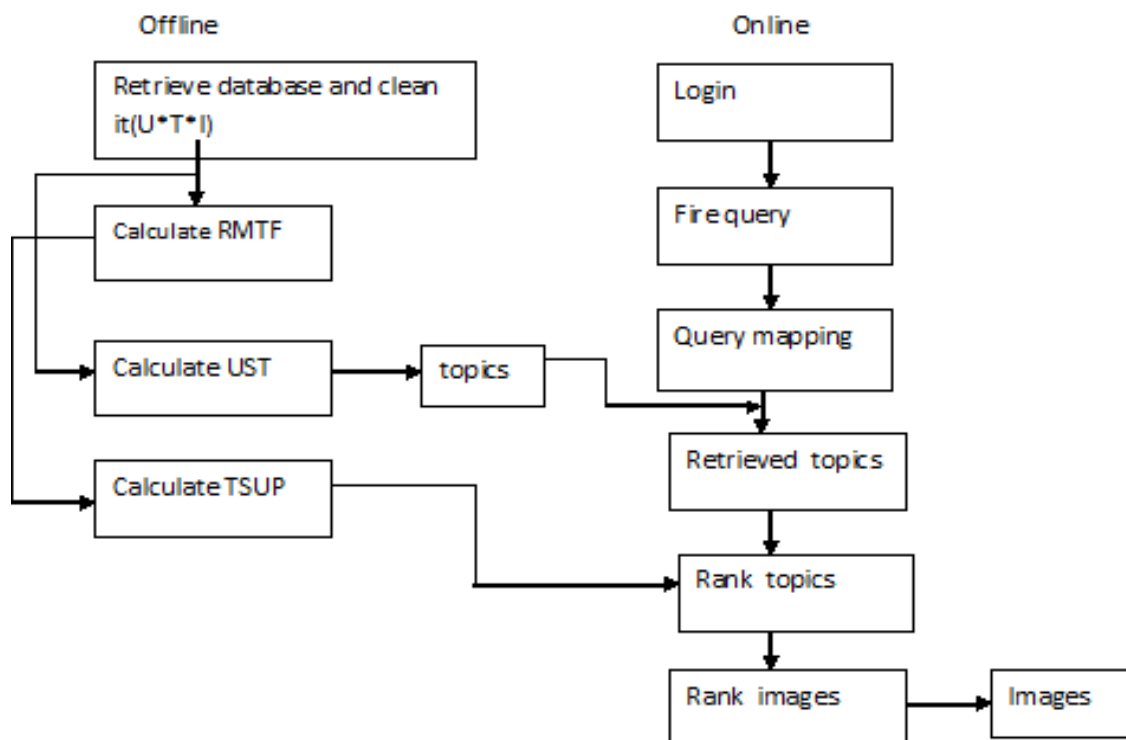


Fig. 2: Architecture of proposed system



Proposed system is worked into two stages i.e. offline and online stages.

**Offline stage:**

1. Ranking Based Multi-correlation tensor Factorization (RMTF)
2. User Specific Topic Modeling (USTM)
3. Topic-Sensitive Users Preferences (TSUP)

**Online stage:**

4. User Specific Query Mapping (USQM)
5. Ranking Based Image Searching.

**1. Ranking Based Multi-correlation tensor Factorization (RMTF)**

When user  $u$  tagged on any particular image id, then that user id, image id, tag named is stored into a database at an offline stage. This database is in the format of ternary interrelationship between users, images and tags. This database is give as an input to RMTF model. The RMTF model calculates user's preferences to assign the tag to a particular image i.e. RMTF provide the users annotation prediction. The tagging data can be viewed as a set of triplets (U I T).RMTF calculates user's preferences by using sigmoid (objective) function sigmoid function retunes values between 0 to 1 that means user preferences lies in between 0 to 1.

**2. User Specific Topic Modeling (USTM)**

After calculating RMTF values, corpus is created for generating topic modeling. Corpus is the folder in which no of folders are created for each user manually. Each folder contains text file for each image and that text file contains tags that user given to that particular image. Corpus is gives as an input to the algorithm.LDA algorithm performed topic modeling. USTM model gives topics for each user; each topic has specific number of relevant terms to each other.

**3. Topic-Sensitive Users Preferences (TSUP)**

USTM model already calculates topics; TSUP calculates topic preferences according to particular user tagging to any image. Each topic having specific number of relevant terms. TSUP calculate preferences for those relevant terms under particular topics. TSUP calculates topic-sensitive user's preferences by using RMTF and USTM model.

**4. User Specific Query Mapping (USQM)**

User fired query  $q$  on search engine then that query  $q$  map from user specific topics. If query  $q$  mapped from USTM then relevant terms of that topic are arranged in ascending order according to topic sensitive user preferences.

**5. Ranking Based Image Searching**

User fired query  $q$  on search engine that the query  $q$  mapped from more than one topic. Then there is need of ranking of those topics according to topic sensitive user preferences. Finally we have obtained highly relevant ranked images .But if query  $q$  does not mapped with any topic in USTM then search engine display normal result like google.

**3.1 Advantages**

1. The proposed system used RMTF model for calculating user annotation prediction. It avoids Common Noisy problem and sever sparsity problem.
2. By using this system we predict the profile of any person.
3. Any person can find the personalized image list easily.
4. User obtained highly ranked images.

**3.2 Algorithm**

The Algorithm to be used in our system:

**3.2.1 LDA algorithm:**

The LDA Algorithm stands for "Latent Dirichlet Allocation Algorithm". This algorithm done the job of user's specific topic generation in the personalized image search

**3.2.2 Steps of algorithm**

- [1] Decide the number of words  $N$  that the document have.
- [2] Go through each document; decide the number of topics  $K$ .

- [3] for D=0 to M do
- [4] for T=0 to K do
- [5] for W=0 to N do
- [6] for i=0 to m do
- [7] for j=1 to n do
- [8] W[i] assigned to first topic.
- [9] Calculate distance between W[i] and W[j];
  - [9.1] if threshold < distance (W[i], W[j])
  - [9.2] W[j] assigned to next topic.
  - [9.3] else W[j] assigned to current topic.
- [10] Repeating the previous step a large number of times until remaining words are allocating in topics.

#### IV. CONCLUSION

Metadata created by users through their everyday activities on social networking site is used to obtain highly relevant images. Ranking Based Multi-correlation tensor factorization is introduced to eliminate the severe sparsity problems appeared in existing system. To find user's topic, LDA (Latent Dirichlet Allocation) algorithm is used. The system introduces two main components to obtain personalized images. First is to calculate user's preferences to assign a tag to the image and second is selection of single keyword query for relevant image searching. Users Sensitive Topics are generated to predict the user's profile. The query mapping or query relevance and topic sensitive user preferences (TSUP) are integrated into final ranked result of relevant images.

#### REFERENCES

- [1]. Jitao Sang, Changchun Xu, Dongyuan Lu "Learn to Personalized Image Search From the Photo Sharing Websites", 2012.
- [2]. D. M. Blei, A. Y. Ng, and M. I. Jordan, "dirichlet allocation" Journal of Machine Learning Research, vol. 3, pp. 993-1022, 2003.
- [3]. F. Qiu and J. Cho, "Automatic identification of user interest for personalized search," in WWW, 2006, pp. 727-736.
- [4]. P.-A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter, "Using odp metadata to Personalize search," in SIGIR, 2005, pp. 178-185.
- [5]. T. G. Kolda and B. W. Bader, "Tensor decompositions and applications" SIAM Review, vol. 51, no. 3, pp. 455-500, 2009.

# Manets: Increasing N-Messages Delivery Probability Using Two-Hop Relay with Erasure Coding

A.Vijayalakshmi<sup>1</sup>, J.Praveena<sup>2</sup>

<sup>1</sup> P.G Student, Tagore Engineering College, Chennai

<sup>2</sup> Asst. Professors, Tagore Engineering College, Chennai

## ABSTRACT:

*The lack of a thorough understanding of the fundamental performance limits in mobile ad hoc networks (MANETs) remains a challenging barrier stunting the commercialization and application of such networks. The proposed system is using a method such as two-hop relay algorithm with erasure coding to increase the message delivery probability of a MANETs. The two-hop relay algorithm with erasure coding used for the message that is erasure coded into multiple frames (coded blocks) at each source node. And also, a finite-state absorbing Markov chain framework is developed to characterize the complicated message delivery process in the challenging MANETs. Based on the developed framework, closed-form expressions are further derived for the message delivery probability under any given message lifetime and message size by adopting the blocking matrix technique where the important issues of interference, medium contention and traffic contention are carefully integrated. To further improve our proposed systems a delivery of n-distinct message is simultaneously send from source to destination.*

**KEYWORDS:** Mobile ad hoc networks, delivery probability, two-hop relay, erasure coding.

## I. INTRODUCTION:

A mobile ad hoc network (MANET) is a peer-to-peer network without any pre-existing infrastructure or centralized administration, which consists of fully self-organized mobile nodes. As it can be rapidly deployed and flexibly reconfigured, the MANET has found many promising applications, such as the disaster relief, emergency response, daily information exchange, etc., and thus becomes an indispensable component among the next generation networks [1], [2]. Since their seminal work in [3], a significant amount of work has been done for a thorough understanding of the delivery delay performance of various routing protocols in MANETs. Zhang et al. in [4] developed an ODE (ordinary differential equations) based framework to analyze the delivery delay of epidemic routing and its variants. Later, Hanbali et al. focused on a multicopy two-hop relay algorithm and explored the impact of packet lifetime (time-to-live TTL) on the packet delivery delay in [5], [6]. Recently, Liu et al. derived closed-form expressions for the packet delivery delay of erasure coding enhanced two-hop relay in [7] and that of group-based two-hop relay in [8]. The delivery delay study [3]–[8], although important and meaningful, can only tell the expected time it takes a routing protocol to deliver a message (or packet) from the source to the destination, i.e., the mean time required to achieve the delivery probability 1, which is actually a simple extreme case of the delivery probability study. Obviously, it is of more interest for network designers to know the corresponding delivery probability under any given message lifetime (or permitted delivery delay). Further notice that in the challenging MANET environment, multiple message replicas are often propagated to improve the delivery performance, where a relay node receiving a message may forward it or carry it for long periods of time, even after its arrival at the destination. Such combination of message replication and long-term storage imposes a severe overhead on the mobile nodes which are usually not only power energy-constrained but also buffer storage-limited. Thus, the message lifetime should be carefully tuned so as to reduce the network resource consumption in terms of buffer occupation and power consumption while simultaneously satisfy the specified delivery performance requirement.

It is noticed that there have been some efforts in literature focusing on the delivery probability study. Panagakis et al. in [9] analytically derived the message delivery probability of two-hop relay under a given time limit by approximating the cumulative distributed function (CDF) of message delivery delay, with the assumption that for any node pair the message can be successfully transmitted whenever they meet each other. In [10], Whitbeck et al. explored the impact of message size, message lifetime and link lifetime on the message delivery ratio (probability) of epidemic routing by treating the intermittently connected mobile networks (ICMNs) as edge-Markovian graphs, where each link (edge) is considered independent and has the same transition probabilities between “up” and “down” status. Later, Krifa et al. in [11] explored the impact of message scheduling and drop policies on the delivery probability performance of epidemic routing, and proposed a distributed scheduling and drop policy based on statistical learning to approximate the optimal performance.

More recently, the optimization issue of message delivery probability under specific energy constraints and message lifetime requirement has also been intensively addressed in the context of delay tolerant networks (DTNs) in which the basic two-hop relay was adopted for packet routing and a wireless link becomes available whenever two nodes meet each other. A common limitation of the available models is that, all these works assumed a single flow (source-destination pair) in their analysis such that all other nodes act as pure relays for this flow. Such models, although simple and easy to use, may neglect an important fact that for the general MANET scenarios, multiple traffic flows may co-exist and a relay node may simultaneously carry messages belonging to multiple flows. Moreover, all these models (no matter for the ICMNs or for the DTNs) focus on a very special MANET scenario, i.e., the sparsely distributed MANET, by assuming that whenever two nodes meet together they can transmit with each other. Obviously, the available models cannot be applied for delivery probability analysis in the general MANETs where the interference and medium contention issues are of significant importance. In this paper, we study the delivery probability performance of two-hop relay MANETs with a careful consideration of the above important issues. The main contributions of this paper are summarized as follows. We focus on the two-hop relay algorithm in MANETs with erasure coding and more general traffic pattern, where the message at each source node is erasure coded into multiple frames (coded blocks). We develop a general finite-state absorbing Markov chain theoretical framework to model the complicated message spreading process in the challenging MANETs, which can also be used to analyze the delivery probability performances under other popular routing protocols, like the epidemic Routing. Based on the Markov chain framework, we further derive closed-form expressions for the corresponding message delivery probability under any given message lifetime and message size by adopting the blocking matrix technique, where the important issues of interference, medium contention and traffic contention in MANETs are carefully incorporated into the analysis. Extensive simulation studies are conducted to validate our theoretical framework, which indicates that the new framework can be used to accurately predict the message delivery probability in MANETs with two-hop relay and erasure coding, and characterizes how the delivery probability varies with the parameters of message size, replication factor and node density there. The rest of this paper is outlined as follows. Section II introduces the system models, the routing protocol and the scheduling scheme considered in the paper. In Section III, we develop a Markov chain theoretical model for the delivery probability analysis under any given message lifetime and message size.

## II. PRELIMINARIES

### 2.1. System Models

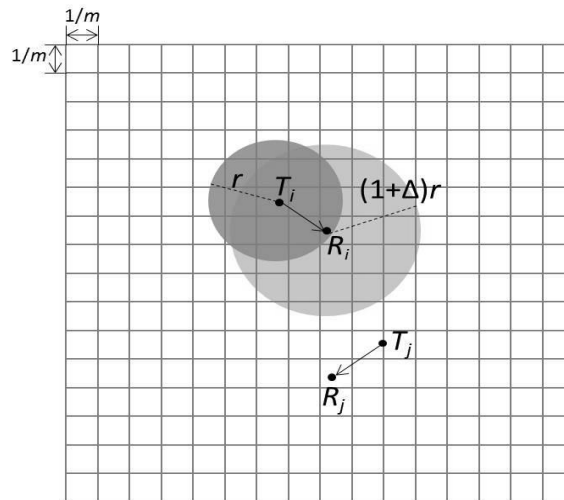
The considered mobile ad hoc network is a unit torus with  $n$  mobile nodes. The torus is evenly divided into  $m \times m$  equal cells (or squares), each cell of side length  $1/m$  as shown in Fig. 1(a). Time is slotted and nodes randomly roam from cell to cell according to the i.i.d. mobility model, which is defined as follows: at time slot  $t = 0$ , a node is initially placed in one of the  $m^2$  cells according to the uniform distribution. The node randomly selects a cell from the  $m^2$  cells with equal probability of  $1/m^2$  independent of other nodes, and moves to the selected cell at time slot  $t = 1$ . The node then repeats this process in every subsequent time slot. One can see that at each time slot, the  $n$  nodes are uniformly and randomly distributed in the  $m^2$  cells. Since the node movements are also independent from time slot to time slot, the nodes are totally reshuffled at each time slot. We employ the protocol model in [25] to address the interference among simultaneous link transmissions. Similar to [23], we assume that each time slot will be allocated only for data transmissions in one hop range. The data transmission model is defined as follows: suppose node  $T_i$  is transmitting to node  $R_i$  at time slot  $t$  as shown in Fig. 1(a), and denote by  $T_{it}$  and  $R_{it}$  the positions of  $T_i$  and  $R_i$ , respectively. According to the protocol model, the data transmission from  $T_i$  to  $R_i$  can be successful if and only if the following two conditions hold for any other simultaneous transmitting node  $T_j$ :

$$|T_{it} - R_{it}| \leq r \quad (1)$$

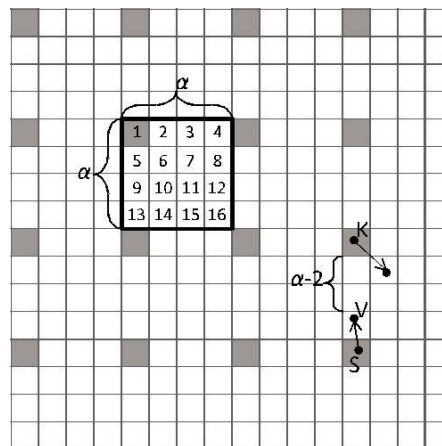
$$|T_{jt} - R_{it}| \geq (1 + \Delta)|T_{it} - R_{it}| \quad (2)$$

Here  $r$  is the transmission range adopted by each node, and  $\Delta > 0$  is a protocol specified factor to represent the guardzone around each receiver. In order to fully characterize the traffic contention issue in MANETs, we consider here the permutation traffic pattern, where each node has a locally generated traffic flow to deliver to its destination and also needs to receive a traffic flow originated from another node. Since there are  $n$  mobile nodes in the network, it is easy to see that there exist in total  $n$  distinct traffic flows. We assume that the local traffic flow at each node has only a single message. Without loss of generality, we focus on a tagged flow hereafter and denote its source and destination by  $S$  and  $D$ , respectively. For the tagged flow, the message generated at the source  $S$  is assumed to have in total  $\omega$  blocks ( $\omega \geq 1$ ), where a single block can be successfully transmitted during a time slot (or meeting duration). We further assume that the message is relevant during  $\tau$  time slots, i.e., the message is labeled with a lifetime of  $\tau$  time slots, and will be dropped from the network if it fails to make itself to the destination  $D$  within  $\tau$  time slots. Remark 1: Note that the node mobility is homogeneous under the i.i.d. model, where during each time slot all node pairs have the same probability to encounter and the  $n$  nodes are uniformly and randomly distributed in the  $m^2$  cells. Such features of homogeneous node meeting and uniform node distribution, although simple and easy to use, are different from other more practical node distributions, like the correlated distribution, the clustered distribution, the home-point distribution, etc.,

where nodes exhibit significant inhomogeneities in spatial distribution over the network area. Remark 2: The main difficulties in extending the i.i.d. model to take into account more complex mobility, such as the random walk model, random waypoint model, random direction model, correlated mobility model, reference point group mobility, etc., are two folds: the first difficulty is to characterize the node meeting process which depends solely on the node mobility, and the second one is to derive the data transmission probability during each node meeting which is related to both the spatial distribution and data transmissions of other nodes.



(a) Illustration of the transmission model.



(b) Cells in a transmission-group with  $\alpha = 4$ .

Figure. 1. Illustration of a cell-partitioned network with  $m = 16$ .

It is notable that in MANETs, even though a node pair meet together in a time slot they may fail to transmit data due to the interference caused by other simultaneous data transmissions in the network. These two difficulties combined together make the analytical modeling of delivery performance in MANETs much more challenging. Actually, the main reason behind adopting the simple i.i.d. model is that it is very helpful to keep the theoretical analysis tractable and thus enables closed-form analytical results to be developed for the message delivery probability in the challenging MANETs.

### 2.2. Two-Hop Relay with Erasure Coding

The two-hop relay, since first proposed by Grossglauer and Tse (2001) in [11], has been extensively explored in literature and proved to be a popular and efficient routing protocol for DTNs. However, its delivery performance remains largely unknown in the general MANET environment, where the interference may create extra delay in the delivery of packets. It is further noticed that extensive simulation delay performance of two-hop relay in DTNs can be improved via incorporating the erasure coding technique.

Therefore, we focus on the two-hop relay with erasure coding in this paper and develop a theoretical framework to analytically study its delivery performance. According to the two-hop relay algorithm with erasure coding [7], [10], for the tagged flow, the message is first erasure coded into  $\omega \cdot \beta$  equal sized frames (or code blocks) after it is locally generated at S, where  $\beta$  is the replication factor. Since each frame is almost the same size as the original block, we assume that it can also be successfully transmitted during a time slot. Any  $(1 + \epsilon) \cdot \omega$  frames can be used to successfully reconstruct the message, where  $\epsilon$  is a small constant and it varies with the adopted erasure coding algorithm. Similar to [7], we ignore the constant  $\epsilon$  here and thus the message can be successfully recovered at the destination D with no less than  $\omega$  frames collected before it expires (i.e., within  $\tau$  time slots). After erasure coding the message into  $\omega \cdot \beta$  frames, the source node S starts to deliver out these frames according to the two-hop relay algorithm [8]. Every time S is selected as the transmitter via the transmission scheduling scheme, it operates as follows:

Step 1: S first checks whether D is in the transmission range. If so, S conducts with D the "source-to-destination" transmission, where a frame is sent directly to D.

Step 2: For the case that D is not in the transmission range of S, if there is no other node in the one-hop neighborhood of S, S remains idle for the time slot; otherwise, S randomly selects a node, say R, from the one-hop neighborhood as the receiver, and flips an unbiased coin; If it is the head, S chooses to perform the "source-to-relay" transmission with R. S initiates a handshake with R to check whether R is carrying a frame received from S. If so, S remains idle for the time slot; otherwise, S sends to R a frame destined for D. Otherwise, S chooses to perform with R the "relay-to-destination" transmission. S first checks whether it is carrying a frame destined for R. If so, S forwards the frame to R; otherwise, S stays idle for the time slot. It is noticed that distinguished from available works which assumed a simple scenario of single traffic flow, we consider in this paper the permutation traffic pattern to fully characterize the traffic contention issue in MANETs. Under such traffic pattern, node may not only carry the frames of its own message, but also simultaneously carry multiple frames originated from other nodes in the network. To simplify the analysis and thus keep the theoretical framework tractable, we assume that each frame will be delivered to at most one relay node and each relay node will carry at most one frame from in S. We consider a single-copy version of the two-hop relay with erasure coding, where S either delivers a frame to D or sends it to a relay node. After sending a frame to a relay node, S retains a copy of the frame as backup; while the relay node will delete the frame from the buffer after forwarding it to D. Therefore, before arriving at D, each frame may have at most two copies in the network, one in the relay node and the other one in S.

### 2.3. Transmission Scheduling

Similar to previous studies we consider a local transmission scenario [7], [8], [9], where a transmitter in a cell can only transmit to receivers in the same cell or other eight adjacent cells (two cells are called adjacent cells if they share a common point). Thus, the transmission range can be accordingly determined as  $r = \sqrt{8}/m$ . It is easy to see that two links can transmit simultaneously if and only if they are sufficiently far away from each other. To avoid collisions among simultaneous transmitting links and support as many simultaneous link transmissions as possible, we adopt here the transmission-group based scheduling scheme [7], [8], [9]. With such a transmission-group definition, all  $m^2$  cells are actually divided into  $\alpha^2$  distinct transmission-groups. If each transmission-group becomes active (i.e., has link transmissions) alternatively, then each cell will also become active every  $\alpha^2$  time slots. As illustrated in Fig. 1(b) for the case  $\alpha = 4$ , there are in total 16 transmission-groups, and all shaded cells belong to the same transmission-group. Setting of Parameter  $\alpha$ : As shown in Fig. 1(b), suppose node S in an active cell is transmitting to node V in a time slot. It is easy to see that another transmitter, say K, in another active cell is at least  $\alpha - 2$  cells away from V. According to the protocol interference model, we should have  $(\alpha - 2) \cdot 1/m \geq (1 + \Delta) \cdot r$  to ensure the successful data reception at V. Notice that  $\alpha \leq m$  and  $r = \sqrt{8}/m$ , then the parameter  $\alpha$  can be determined as

$$\alpha = \min\{[(1 + \Delta) \sqrt{8} + 2] m\} \quad (3)$$

It is noticed that with the setting of  $\alpha = m$ , all network cells are divided into  $m^2$  distinct transmission-groups, with each transmission-group containing a single cell. Therefore, the network can support only one active transmitter-receiver pair during each time slot. For the transmission-group based scheduling scheme, a node is assumed to be able to obtain the cell id where it resides at the beginning of each time slot. Actually, for a given network cell partition, such hypotheses can be satisfied by adopting the global positioning system (GPS) or some node localization schemes. Therefore, after obtaining the cell id, a node can easily judge whether it is inside an active cell or not for the current time slot, and then the nodes in an active cell can compete to become the transmitter via a distributed coordination function (DCF)-style mechanism.

Remark 3: The transmission-group based scheduling scheme has the following two advantageous features: firstly, it is fully distributed and thus it can be implemented without any centralized management; secondly, it enables closed-form expressions to be derived for the transmission probability under the two-hop relay during each time slot. It is also noticed that in (1) we derive  $\sqrt{\alpha}$  according to the possible maximum distance (i.e.,  $r = \sqrt{8/m}$ ) between a transmitter-receiver pair in two adjacent cells. However, one can see that the distance between a transmitter-receiver pair selected for each active cell may be less than  $\sqrt{8/m}$  with high probability during each time slot. Consequently, the scheduling scheme may unavoidably result in an inefficient spatial reuse due to the fixed setting of  $\alpha$ .

### III. MESSAGE DELIVERY PROBABILITY

In this section, we first introduce some basic probabilities under the two hop relay with erasure coding, develop the Markov chain theoretical framework, and then proceed to derive the message delivery probability.

#### 3.1. Markov Chain Framework

For the tagged flow, as the message generated at the source node S is erasure coded into  $\omega \cdot \beta$  frames and is relevant only in  $\tau$  time slots, the destination node D needs to collect at least  $\omega$  frames within  $\tau$  time slots so as to successfully recover the message. If we denote by  $(j, k)$  a general transient state during the message delivery process that S is delivering the  $j$ th frame and D has already received  $k$  distinct frames, and further denote by  $(*, k)$  a transient state that S has already finished dispatching all  $\omega \cdot \beta$  frames while D has only received  $k$  of them,  $1 \leq j \leq \omega \cdot \beta$ ,  $0 \leq k < \omega$ , then we can characterize the message delivery process with a finite-state absorbing Markov chain. Specifically, if the tagged flow is in state  $(j, k)$  at the current time slot, only one of the following four transition cases illustrated in Fig. 2 may happen in the next time slot.

SR Case: "source-to-relay" transmission only, i.e., S successfully delivers the  $j$ th frame to a new relay node while none of the relays delivers a frame to D. As shown in Fig. 2(a) that under such a transition case, the state  $(j, k)$  may transit to two different neighboring states depending on the current frame index  $j$ .

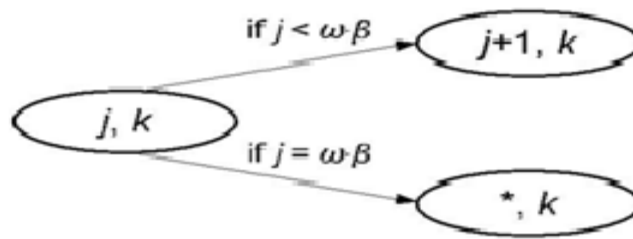


Figure 3.1 SR Case

RD Case: "relay-to-destination" transmission only, i.e., some relay node successfully delivers a frame to D while S fails to deliver out the  $j$ th frame to a new relay node. As shown in Fig. 2(b) that there is only one target state  $(j, k + 1)$  under the RD case.



Figure 3.1 RD Case

SR+RD Case: both "source-to-relay" and "relay-to-destination" transmissions, i.e., these two transmissions happen simultaneously. We can see from Fig. 2(c) that depending on the value of  $j$  there are two possible target states under the SR+RD case.

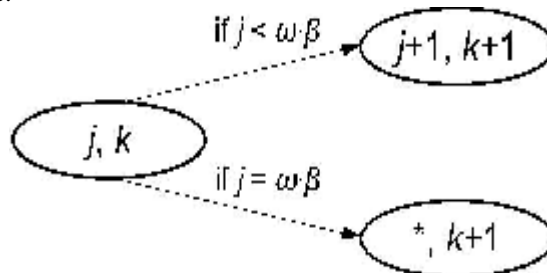


Figure 3.1 SR+RD case

SD Case : “source-to-destination” transmission only, i.e., S successfully delivers a frame to D. As shown in Fig. 2(d) that under the SD case, the state (j, k) may transit to (j +1, k +1) or (\* , k +1), similar to that under the SR+RD case.

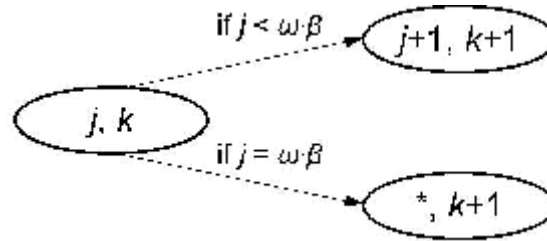


Figure 3.1 SD case

### 3.2. Derivations of delivery probability $\phi(\omega, \beta, \tau)$

Before deriving the message delivery probability, we first node S which is further erasure coded into  $\omega \cdot \beta$  frames, the delivery delay of the message and the timeslot when the destination node D receives  $\omega$  distinct frames of the message. For the tagged flow, if we denote by  $T_d$  the message delivery delay and denote by  $\phi(\omega, \beta, \tau)$  the message delivery probability under the message lifetime constraint  $\tau$ , then we have

$$\phi(\omega, \beta, \tau) = \Pr(T_d \leq \tau) = \sum_{t=1}^{\tau} \Pr(T_d = t). \quad (4)$$

Based on the Markov chain framework, now we are ready to derive  $\phi(\omega, \beta, \tau)$ . All  $\delta$  transient states in the Markov chain are arranged into  $\omega$  rows. We number these transient states sequentially  $1, 2, \dots, \delta$  in a left to right and top to bottom way. For these transient states, if we let  $q_{ij}$  denote the transition probability from state  $i$  to state  $j$ , then we can define a matrix  $Q = (q_{ij})_{\delta \times \delta}$  of transition probabilities among  $\delta$  transient states there. Similarly, if we let  $b_i$  denote the one step transition probability from state  $i$  to the absorbing state A, then we can also define a vector  $B = (b_i)_{\delta \times 1}$  representing the transition probabilities from  $\delta$  transient states to state A. Notice that  $\Pr(T_d = t)$  in (4) denotes the probability that the  $\omega$ th frame arrives at the destination D by the end of the  $t$ th time slot, i.e., the probability that the Markov chain gets absorbed by the end of the  $t$ th time slot. Given that the Markov chain starts from the first state, i.e., state  $(1, 0)$ , according to the Markov chain theory, then we have

$$\Pr(T_d = t) = \sum_{i=1}^{\delta} q_{1i}^{t-1} \cdot b_i \quad (5)$$

Then combining with the fact of previous equation is actually the  $ij$ -entry of the matrix  $Q^t$  can be further transformed as

$$\Pr(T_d = t) = e \cdot Q^{t-1} \cdot B \quad (6)$$

where  $e = (1, 0, \dots, 0)$ .

Substituting (14) into (12), then we have

$$\begin{aligned} \phi(\omega, \beta, \tau) &= e \cdot Q^{t-1} \cdot B \\ &= e \cdot (I - Q)^{-1} \cdot (I - Q^{\tau}) \cdot B \\ &= e \cdot N \cdot (I - Q^{\tau}) \cdot B \end{aligned} \quad (7)$$

where  $I$  is the identity matrix, and  $N = (I - Q)^{-1}$  is the fundamental matrix of the Markov chain.

## IV. RESULTS AND DISCUSSION

In this paper, we have investigated the message delivery probability in two-hop relay MANETs with erasure coding. A general Markov chain theoretical framework was developed to characterize the message delivery process, which can also be used to analyze the delivery probability performances under other popular routing protocols. Based on the new theoretical framework, closed-form expressions were further derived for the delivery probability under any given message lifetime and message size. As verified by extensive simulation studies, our framework can be used to efficiently model the message delivery process and thus accurately characterize the delivery probability performance there. Our results indicate that for a two-hop relay MANET with erasure coding, there exists a limiting performance for the delivery probability, which is determined only by the control parameters of message size  $\omega$  and message lifetime  $\tau$ . Another interesting finding of our work is that the considered MANETs actually exhibit very similar behaviors in terms of delivery probability under different node mobility models, like the i.i.d., random walk and random waypoint.



## V. CONCLUSION

This project can be design a future network in terms of determining a suitable message lifetime, so as to minimize the per node buffer occupation and power consumption while simultaneously meet the specified delivery performance requirement. Note that the theoretical framework and closed-form results developed in this paper only hold for the simple scenario that each node has only a single message to deliver to its destination, and it chooses to conduct a "source-to-relay" or "relay-to-destination" transmission in a probabilistic fashion. Therefore, one future work is to further explore the delivery probability of two-hop relay with erasure coding in a more general scenario, where each node may need to simultaneously deliver  $k$  distinct messages, and it conducts the "source-to-relay" or "relay-to-destination" transmission in the best-effort fashion so as to take the full advantage of each transmission opportunity. Another interesting future direction is to extend the theoretical models in this paper to analytically derive the optimum setting of  $n$  (i.e.,  $n^*$ ) to achieve the maximum message delivery probability for a given relay scheme setting of  $(\omega, \beta, \tau)$ , or to formally determine the asymptotic (limiting) delivery probability for any specified control parameters of  $\omega$  and  $\tau$ . This project implemented a Mobile Ad-hoc Network model for the compute message delivery probability. The network is functioned based on effective transmission group scheduling scheme, two-hop relay mechanism and Tornado encoding scheme. The proposed network model sends the multiple selected messages in effective way. Also, a dynamic memory management mechanism is utilized to manage the sending messages. Finally, a modified Markov chain model is utilized to compute the message delivery probability of whole selected message. In future, This project can be design a future network in terms of determining a suitable message lifetime, so as to minimize the per node buffer occupation and power consumption while simultaneously meet the specified delivery performance requirement.

## REFERENCES

- [1] Beyond Shannon: The Quest for Fundamental Performance Limits of Wireless Ad Hoc Networks", Andrea Goldsmith, Stanford University, Michelle Effros, Caltech Ralf Koetter, Technical University of Munich Muriel Médard, Asu Ozdaglar, and Lizhong Zheng, MIT may 2011.
- [2] Optimal Delay-Throughput Trade-Offs in Mobile Ad Hoc Networks", Lei Ying, Sichao Yang, and R. Srikant Coordinated Science Lab and Department of Electrical and Computer Engineering University of Illinois at Urbana-Champaign, 2007.
- [3] Optimal Control in Two-Hop Relay Routing", Eitan Altman, Tamer Basar, and Francesco De Pellegrini, 2004.
- [4] Rethinking Information Theory for Mobile Ad Hoc Networks", Jeff Andrews, Nihar Jindal, Martin Haenggi, Randy Berry, Syed Jafar, Dongning Guo, Sanjay Shakkottai, Robert Heath, Michael Neely, Steven Weber, Aylin Yener, 2007.
- [5] Routing in Delay/Disruption Tolerant Networks: A Taxonomy, Survey and Challenges", Yue Cao and Zhili Sun, Member, IEEE, 2012
- [6] Annotating Structured Data of the Deep Web", Hai He<sup>1</sup>, Weiyi Meng<sup>1</sup>, Hongkun Zhao<sup>1</sup>, and Clement Yu Department of Computer Science SUNY at Binghamton, NY 13902, USA
- [7] A. A. Hanbali, A. A. Kherani, and P. Nain, "Simple models for the performance evaluation of a class of two-hop relay protocols," in Proc. 2007 IFIP Netw. Conf.
- [8] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Delay and capacity in ad hoc mobile networks with  $f$ -cast relay algorithms," IEEE Trans. Wireless Commun., vol. 10, no. 8, pp. 2738–2751, Aug. 2011.
- [9] S. R. Kulkarni and P. Viswanath, "A deterministic approach to through-put scaling in wireless networks," IEEE Trans. Inf. Theory, vol. 50, no. 6, pp. 1041–1049, June 2004.
- [10] Y. Wang, S. Jain, M. Martonosi, and K. Fall, "Erasure-coding based routing for opportunistic networks," in Proc. 2005 WDTN Conf.
- [11] M. Grossglauser and D. N. Tse, "Mobility increases the capacity of ad hoc wireless networks," in Proc. 2001 IEEE INFOCOM.

## Survey Paper on Virtualized Cloud Based IPTV System

Jyoti P. Hase<sup>1</sup>, Prof. R.L.Paikrao<sup>2</sup>

<sup>1</sup>M.E.ComputerAVCOE, Sangamner, India

<sup>2</sup>Dept. of Computer Engg.AVCOE, Sangamner, India

### ABSTRACT

*IPTV delivers the TV content over an internet Protocol infrastructure. Virtualized cloud-based services will benefit of stastical multiplexing across applications to yield important cost savings to the operator the cloud based IPTV provide lower a provider's costs of real-time IPTV services through a virtualized IPTV architecture and through intelligent time shifting of service delivery. It takes advantage of the differences in the deadlines associated with Live TV versus Video-on-Demand (VoD) to effectively multiplex these services. However, achieving similar advantages with period of time services are often a challenge. For construct the problem as an optimization formulation that uses a generic cost function. e.g., minimum-maximum, concave and convex functions to reflect the different cost operation. We are going to study The time shifting solution to this formulation gives the number of servers needed at different time instants to support these services. a simple mechanism for time-shifting scheduled jobs in a simulator and study the reduction in server load using real traces from an operational IPTV network. The cloud based IPTV results show that able to reduce the load by 24%.*

### I. INTRODUCTION

As IP-based video delivery becomes more famous, the demands places upon the service provider's resources have dramatically accrued. Service suppliers usually provision for the height demands of every service across the Provider population. However, provisioning for peak demand leaves resources underneath used in any respect alternative periods. This is often significantly evident with Instant Channel change (ICC) requests in IPTV[1]. In IPTV, Live TV is often multicast from servers exploitation IP Multicast, with one cluster per TV channel. Video-on-Demand (VoD) is additionally supported by the service Providers with each request being serve by a server serving a unicast stream When users amendment channels whereas observation live TV, we need to give extra practicality to so the channel change takes impact quickly[1]. For every channel amendment, the user has to be part of the multicast cluster related to the channel, and expect enough information to be buffered before the video is displayed; this may take your time. As a result, there are many makes an attempt to support instant channel amendment by mitigating the user perceived channel shift latency. With the typical Instant Channel Change enforced on IPTV systems, the contents is delivered at happen more quickly rate using a unicast stream from the server. The playout buffer is stuffed quickly, and so keeps switching latency little. Once the playout buffer is stuffed up to the playout purpose, the set top box activity back to receiving the multicast stream. Our Aim is in this paper is to take advantage of the distinct workloads of the various IPTV services to higher utilize the deployed servers. It offers opportunities for the service provider to deliver the VoD content in anticipation and potentially out of order, taking advantage of the buffering available at the receivers . Virtualization offers us the flexibility to share the server resources across these services & use a cloud computing infrastructure with virtualization is shifting the resources dynamically in real time to handle the instant channel change workload, b) to be ready to anticipate the amendment within the work before time and preload VoD content on Set Top Boxes, thereby facilitate the shifting of resources from Video on Demand to Instant Channel Change during the bursts and c) to solve a general cost problem of optimization formulation without having to meticulously model every and each parameter setting in an exceedingly data center to facilitate this shifting resources. In a virtualized surroundings, Instant channel change is managed by a group of VMs typically, alternative VMs would be created to handle Video on Demand requests . With the power to spawn VMs quickly [4] .

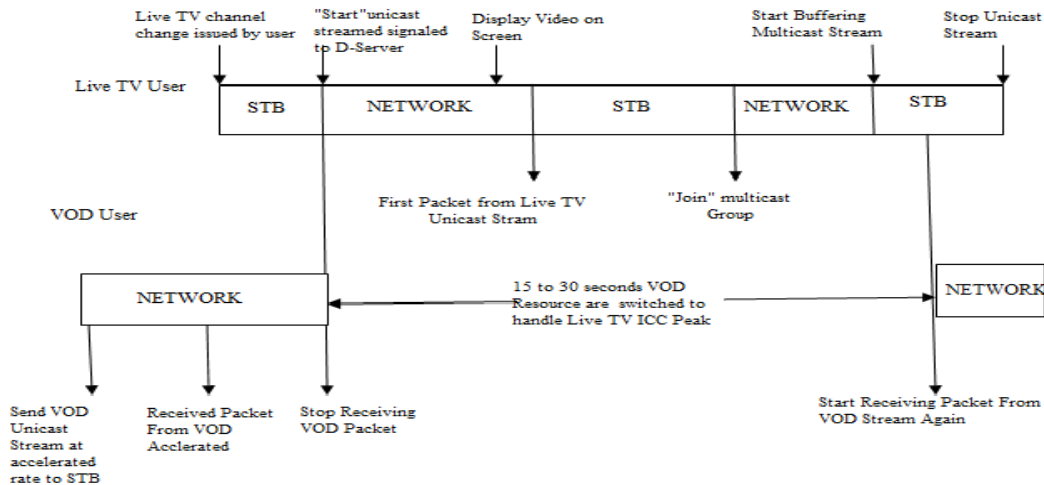


Fig1. Live TV ICC and VoD packet buffering timeline

the server will be shift (VMs) from Video On Demand to handle the Instant Channel Change demand in a matter of some seconds[5] Note that by having the ability to predict the Instant channel Change bursts channel modification behavior are often foretold from historic logs as a results of Live Television show timings. The channel changes sometimes happens each hour. In anticipation of the Instant Channel Change load.

## II. BACKGROUND

### 2.1 Architecture of IPTV System

The IPTV architecture we are implementing this Architecture on. cloud network.fig 2 shows the IPTV Architecture and system components for IPTV systems. The distribution network consists of video servers for every metropolitan space connected through the metro network of internet protocol routers and optical networks to the access network[2]. The access network generally includes a tree-like distribution network with copper or fiber (newer environments) property to the house. In associate example nationwide distribution setting, content is received from the point wherever it's non inheritable over an Internet Protocol backbone to the IPTV head-end for the metropolitan space at a Video Hub Office (VHO). From now, the video is distributed to subscriber homes within the metro or cloud Network.

#### 1 Content sources & D-Server:

Video content is received from content suppliers either live or from storage (for Video on Demand). The content is buffered at Distribution Servers (D-Server) within the Video hub Office (VHO). A different D-server can be used for each channel; all D-servers share the link to the Video Hub Office, and on average one D-server's input output capacity is of the order of ~100s of Mbps

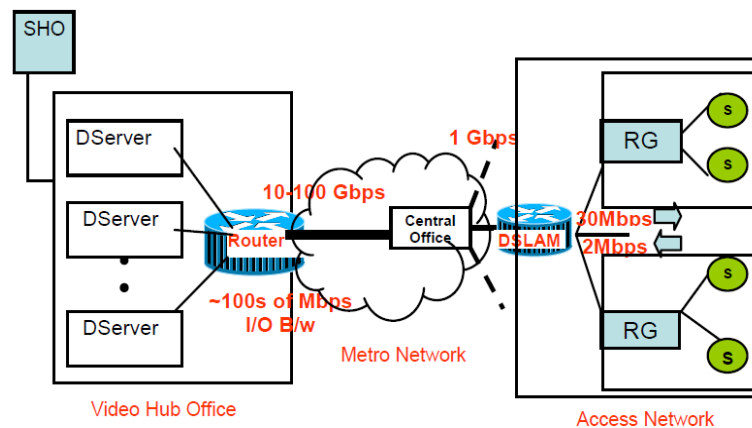


Fig. 2 IPTV Architecture & component [2]

2. Metro Network or cloud network: it connects the VHO to variety of central offices (CO). The Metro network is usually Associate in Nursing optical network with vital capability. Downstream of the Central Office are many Digital Subscriber Line Access Multiplexers (DSLAMs).

3. Customer access link: Delivery of IPTV over the “last mile” is also provided over existing loop plant to homes victimization higher-speed phone line technologies like VDSL2. Service suppliers could use a mix of Fiber-to-the-Node (FTTN) and phone line technologies or implement direct Fiber-to-the-Home (FTTH) access.

4. client premises equipment (CPE):CPE includes the broadband network termination (B-NT) and a Residential gateway (RG). The RG is generally an Internet protocol router and is the demarcation point between the service supplier and also the home network.

5. IPTV Client: The IPTV consumer (e.g., Set top Box (STB)) terminates IPTV traffic at the designer premises.

## 2.2 Unicast Instant Channel Change (ICC)

We are studying ICC For channel Changing Operation Fig 3 shows the working of ICC whenever user wants to change the channel. When a client connects to the D-server, the Dserver begins unicasting data to the client, starting from an Iframe in its buffer. The D-Server bursts the data at a higher rate than the nominal video bitrate rate. Given this higher unicast rate, the set-top box (STB) buffer fills up faster than the nominal rate at which the multicast stream is transmitted[2]. A multicast “join” is issued by the client after sufficient data is buffered in the playout buffer of the STB so that the buffer does not under-run by the time the multicast join completes and the subsequent multicast stream is received. The unicast stream is stopped once the playout buffer is filled to the desired level. Once the multicast join is successful, the client can start displaying video received from the multicast stream.

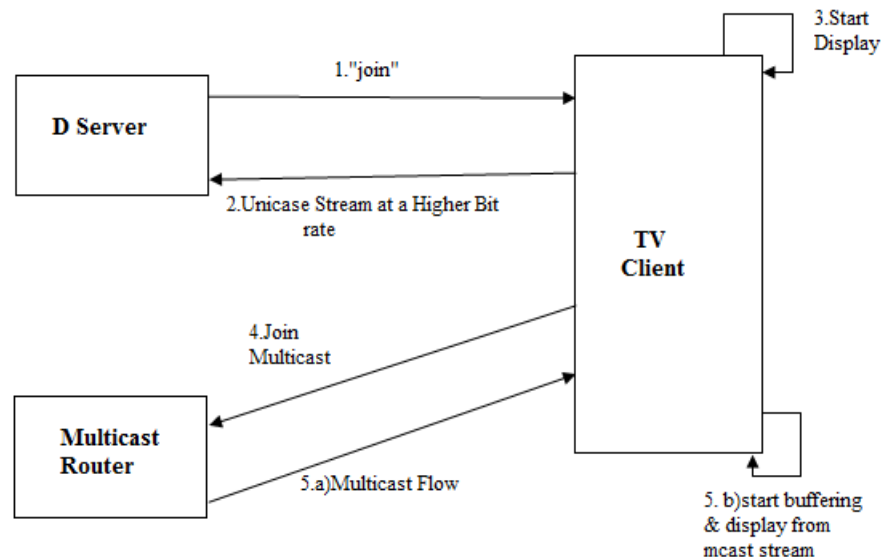


Fig. 3 Unicast Instant channel change scheme

We consider various cost functions  $C(x_1, x_2, \dots, x_T)$ , evaluate the optimal server resources needed, and study the impact of each cost function on the optimal solution.

## 2.3 Multicast ICC Scheme

We studying a secondary lower-bandwidth channel change stream corresponding to each channel at the D-server. This stream will consist of I-frames only Therefore, each channel will now add another IP multicast group called the secondary ICC multicast group that transmits only the I frames for each channel. The secondary ICC channel change stream is offset by a short time interval so that the secondary stream is delayed in time relative to the primary high quality multicast stream. This enables the client to display the (less than full motion) video of the new channel that the user switches to, while allowing the play-out buffer to be filled with the primary multicast stream.

Figure 4 shows the mechanism of the Multicast ICC process:

1. The channel change request for a particular channel results in a multicast join request being issued by the client. A join is issued for both the primary multicast stream as well as the secondary ICC multicast group for the channel change stream obtained by extracting the I frames from the primary stream. With IP multicast, the join progresses as far up the distribution tree as necessary, until it hits an "on-tree" node.
2. The D-server transmits both the secondary channel change stream as well as the primary multicast channel if there is an outstanding channel change event for that channel.

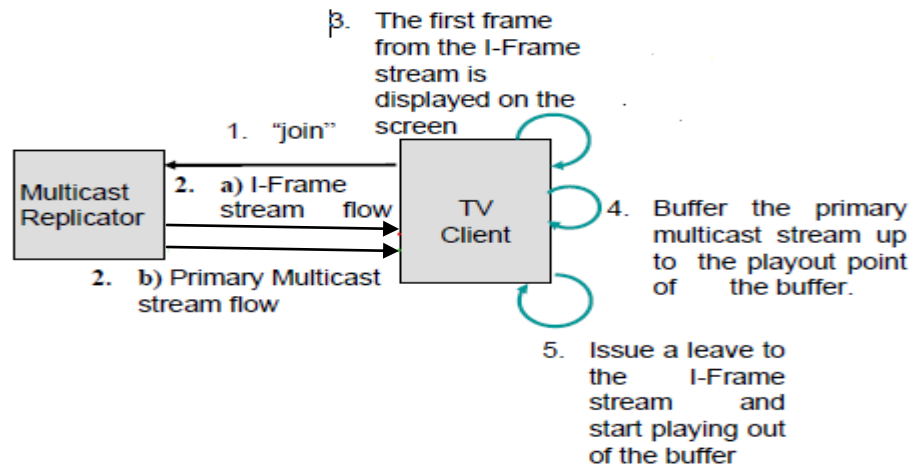


Fig.4 Multicast ICC channel Scheme

#### 2.4 Mathematical function

We investigate linear, Concave and convex. With convex functions, the price will increase slowly initially and later grows quicker. For concave functions, the cost will increase quickly at first and so flattens out, indicating some extent of decreasing unit prices (e.g., slab or tiered pricing). Minimizing a convex cost function leads to averaging the amount of servers. Minimizing a concave cost results in finding the external points off from the most to reducing the cost. This might result in the system holding back the requests till simply before their deadline and serving them in an exceedingly burst, to induce the advantage of a lower cost due to concave cost function (e.g., slab pricing). The concave optimization problem is so optimally solved by finding boundary points within the server-capacity region of the answer area.

We study various cost functions  $C(x_1, x_2, \dots, x_T)$ , evaluate the optimal server resources needed, and study the impact of each cost function on the optimal solution.

We think about the subsequent cost functions:

1) **Linear Cost:**  $C(x_1, x_2, \dots, x_T) = \sum_{i=0}^T x_i$ .

the case where we incur a cost that's proportional to the total number of servers required across all times.

2) **Convex Separable Cost:**  $C(x_1, x_2, \dots, x_T) = \sum_{i=0}^T C(x_i)$ .

Where  $C(x_i)$  is a convex function. In case when a data center sees an increasing per unit cost as the number of servers needed to grow.

We consider two examples of  $C(x_i)$ , the component cost function. The 1st is the exponential function,  $C(x_i) = \exp(x_i)$ .

The 2<sup>nd</sup> is a piecewise linear function of the form  $C(x_i) = x_i + c(x_i - K)^+$  where  $c, K \geq 0$ . This component cost function has per-server cost of unit when  $x_i \leq K$ , and per-server cost of  $1 + c$  thereafter.

3) **Concave Separable Cost:**  $C(x_1, x_2, \dots, x_T) = \sum_{i=0}^T C(x_i)$ ,

with component cost  $C(x_i)$  a concave function. This part value perform has per-server cost diminishes as the number of servers grows.

4) **Maximum Cost:**  $C(x_1, x_2, \dots, x_T) = \max_{i=1}^T x_i$

This cost function penalizes the highest capacity that which will be required to use the incoming sequences of requests

#### 2.5 Advantages of cloud Based IPTV System

In Virtualized cloud Based system there is more workload so reducing workload we are designing the Group Of Picture is to secure the transmission of huge amount of data over the unreliable channel e.g. internet,

wireless channels etc .Software will hide text document in container video file[3]. Appropriate container video file must be selected such that it would be feasible to transmit over internet as such video file requires large bandwidth The software must hide large amount of data The software must encrypt the data before hiding it into container media file Message hiding should be such that hidden message must intercept table so the system gives following **Advantages** -

### 1.Live TV Controller

Rewind, pause, play while watching television broadcast due to that it is called triple play.

### 2. Video-On-Demand

IPTV technology is bringing video-on-demand (VoD) to television, which permits a customer to browse an online program or film catalog, to watch trailers and to then select a selected recording.

### 3. Interactivity

An IP-based platform additionally permits vital opportunities to form the TV viewing expertise a lot of interactive and customized ,The provider program guides that enables viewers to go looking for content by title or actors name, or a picture-in-picture practicality that enables them to channel surf without departure the program observation.

### 4. Economics

However, because video streams need a high bit rate for for much longer periods of your time, the expenditures to support high amounts of video traffic are abundant bigger.

### 5. IPTV primarily based Converged Service

Another advantage of an IP-based network is that the chance for integration and convergence. this chance is amplified once exploitation IMS-based solutions.[6] Con- verged services implies interaction of existing services in an exceedingly seamless manner to form new worth side services. One example is on-screen display, obtaining display on a TV and also the ability to handle it (send it to voice mail, etc.). IP-based services can help to modify efforts to supply customers anytime-anywhere access to content over their televisions, Personal Computers and cell phones, and to integrated services and content to them along at intervals businesses and establishments, IPTV eliminates the requirement to run a parallel infrastructure to deliver live and hold on video services.

### 6. Ease of installation and operation.

### 7. Competitive pricing.

## III. CONCLUSION

We studied however IPTV service suppliers will leverage a virtualized cloud in-frastructure and intelligent time-shifting of load to higher utilize deployed resources. Time shifting reduces the workload as well Instant Channel Change and VoD delivery as examples. we studied that we will benefit of the distinction in workloads of IPTV services to schedule them befittingly on virtualized infrastructure. formulate to developed as a general optimization problem and computed the quantity of servers needed in line with a generic value operate. We studied multiple forms for the value operate (e.g., min-max, convex and concave) and solved for the best variety of servers that are needed to support these services while not missing any deadlines.

## REFERENCES

- [1] "Optimizing Cloud Resources for Delivering IPTV Services through Virtualization", Vaneet Aggarwal, Vijay Gopalakrishnan, Rittwik Jana, K. K. Ramakrishnan, Vinay A. Vaishampayan AT&T Labs – Research, 180 Park Ave, Florham Park, NJ, 07932
- [2] D.Banodkar, K.K. Ramakrishnan, S. Kalyanaraman, A. Gerber, and O. Spatscheck, "Multicast instant channel change in IPTV system," in Proceedings of IEEE COMSWARE, January 2008.
- [3] "Microsoft tv: Iptv edition," [http:// www. microsoft. com/tv/IPTVEdition .msp](http://www.microsoft.com/tv/IPTVEdition.msp).
- [4] H. A. Lagar-Cavilla, J. A. Whitney, A. Scannell, R. B. P. Patchin, S.M.Rumble, E. de Lara, M. Brudno, and M. Satyanarayanan, "SnowFlock: Virtual Machine Cloning as a First Class Cloud Primitive," ACM Transactions on Computer Systems (TOCS), 2011.
- [5] V. Aggarwal, V. Gopalakrishnan, R. Jana, K. K. Ramakrishnan, and V. Vaishampayan, "Exploiting Virtualization for Delivering Cloud-based IPTV Services," in Proc. of IEEE INFOCOM (mini-conference), Shanghai, April 2011.
- [6] R. Urgaonkar, U. Kozat, K. Igarashi, and M. J. Neely, "Dynamic resource allocation and power management in virtualized data centers," in Proceedings of IEEE IFIP NOMS, March 2010.

# The Search of New Issues in the Detection of Near-duplicated Documents

Hasan Naderi (PHD)<sup>1</sup>, Narges salehpour<sup>2</sup>, Mohammad Nazari farokhi<sup>3</sup>,  
Behzad Hosseini chegeni<sup>4</sup>

<sup>1</sup>Department of Computer  
Engineering(IUST)

<sup>2</sup>Department of Computer  
Engineering(IAU)

<sup>3</sup>Department of Computer  
Engineering(IAU)

<sup>4</sup>Department of Computer  
Engineering(IAU)

## ABSTRACT:

Identifying the same document is the task of near-duplicate detection. Among the near-duplicate detection algorithms, the fingerprinting algorithm is taken into consideration using in analysis, plagiarism, repair and maintenance of social softwares. The idea of using fingerprints is in order to identifying duplicated material like cryptographic hash functions which are secure against destructive attacks. These functions serve as high-quality fingerprinting functions. Cryptographic hash algorithms are including MD5 and SHA1 that have been widely applied in the file system. In this paper, using available heuristic algorithms in near-duplicate detection, a set of similar pair document are placed in a certain threshold, an each set is indentified according to being near- duplicate. Furthermore, comparing document is performed by fingerprinting algorithm, and finally, the total value is calculated using the standard method.

**Keywords:** Near-Duplicate Detection, Fingerprinting, Similarity, Heuristics, Shingle, Hash, Random

## I. INTRODUCTION

Fingerprinting and watermarking in relational database, is used in order to protect copyright, tamper detection, traitor tracing and keep data integrity. Since the web is full of copied content so, detection duplicate documents is difficult. One of the methods that is used for detecting repeated, is the last editing which is done on the document and, the exact solution can be found by fingerprinting method, and its duplication is detectable. Near-duplicate is also used in many documents of the World Wide Web. Among negative effects on Web search engines can name storage space, waste indicators, and results of the same writings cluttered and plagiarism which are the results of repetition. Aside from the deliberate repetition of content, copying is done randomly in companies, universities or different departments that store documents. A Possible solution for solving the problems mentioned is that reliable knowledge to near-duplicate must be existed.

$$\varphi(D, dq) \geq 1 - \epsilon \text{ with } 0 < \epsilon < 1$$

So that  $\varphi$  in the interval  $[0, 1]$  is considered as a similarity function. However, to achieve comprehensiveness each pair of documents must be analyzed. According to that is  $Dq \subset d$ , the comparison of  $\varphi(d, d)$  is needed [1]. The main idea of the similarity between two document  $dq, d$  is using of fingerprinting that fingerprint  $Fd$  is a set of numbers  $k$  from  $d$ , on the other words, if we have two document fingerprinting by  $Fd$  and  $Fdq$ , and  $dq, d$  have near-duplication, the similarity of two documents is estimated as follows via the Jaccard coefficient. (The numbers  $k$  that has subscription can be  $k \leq k$ ):

$$\varphi(D, dq) \geq 1 - \epsilon \text{ with } 0 < \epsilon < 1 \text{ Is close to } \rightarrow \frac{|Fd \cap Fdq|}{|Fd \cup Fdq|}$$

The meaning of fingerprints community is that all documents are in  $D$ , in fact  $X \rightarrow \mu(X), Fd \rightarrow D$ . The inverted file index must be  $X \in Fd$ . In other words  $X$ , is included all documents, and also  $\mu(X)$  is the Postlist of  $X$ . To document  $dq$  with fingerprinting  $Fdq$ , now the collection  $Dq \in D$  includes documents that contain at least  $K$ , Postlist but provided that  $X \in Fdq$  and  $\mu(X)$  is existed. If  $(Dq)$  includes all the documents,  $Fdq$  shares fingerprints for the minimum number  $K$ , according to the  $(Dq)$  is as a heuristic approximation of the function  $Dq$  recovery depends on the construction of fingerprinting, it is calculated as follows:

$$\text{req} = \frac{\delta q \cap Dq}{Dq}, \text{Pre} = \frac{\delta q \cap Dq}{Dq}$$

## II. MATERIALS AND METHODS

### 2.1. Construction of fingerprinting

A piece of document into a sequence of  $n$  words can be cascaded as  $Cd$ .  $Cd$  is a collection of all different patches of documents which its size measures form  $|d| - N$  and in time  $o(|d|)$  is measurable Provided that  $c \in cd$  and  $c$  are one

dimension qualifier with a non-zero weight. Three steps in the construction of fingerprinting must be understood [1]:

#### 2.1.1. Dimensions reducing with mapping are realized

This algorithm selects the dimensions of the previous dimensions so that,  $d$  is unchangeable and  $\bar{d}$  is for reducing the vector. The algorithm is modified in  $d$  mode and additional information may be deleted until it would be possible.

#### 2.1.2. Counting vector (quantization) of $d$ elements

It contains a finite number of integers  $d$ .

#### 2.1.3. Calculation coding of one or more $d$ " code which eventually led to the $d$ be fingerprinted

Fingerprinting algorithms primarily are different, and are used in the reduced dimensions method [2].

Figure 1 shows the organization of fingerprinting algorithms construction method.

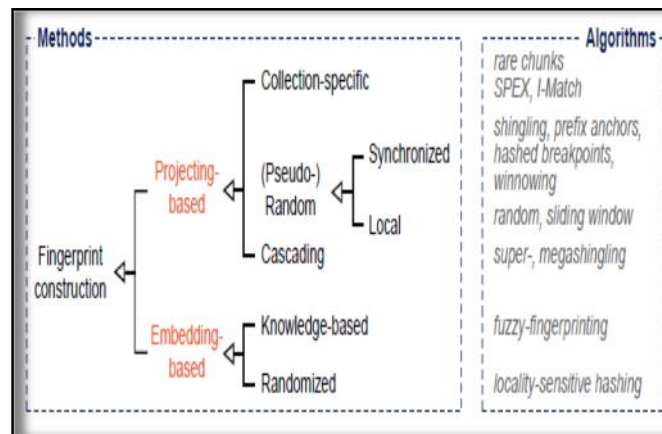


Figure 1: Structure of fingerprints [1].

### 2.2. Mapping to decrease the dimensions

Fingerprinting for  $Fd$  as follows:

$$F d = \{h(c) \mid c \in cd \text{ and } \sigma(c) = \text{true}\}$$

$\sigma$  denotes the heuristics selection to reduce the dimensions that are correct and the condition is satisfied in it, and it would be realized as a piece of document with special characteristics. In this algorithm,  $h$  represents a hash function like Rabins and MD5 algorithm that hash function acts as quantization. Purpose of  $\sigma$  is selecting pieces of document for fingerprints to identify near duplicated documents that are reliable and suitable [1].

### 2.3. Embedded dimension reduction

The fingerprinting is based on embedding  $Fd$  which normally made to a document called similarity hashing that unlike function hash is standard. Its goal is to minimize the number of hash collisions.

The hash function is  $h\phi: D \rightarrow U, U \subset N$  and also  $\phi(d, dq) \geq 1-\epsilon$  if  $(d, dq) \in D$ . In order to make fingerprinting  $Fd$  in  $d$  document, a limited number of  $k$  models that have used the function  $h$ , must form:  $Fd = \{h\phi^{(i)}(d) \mid i \in \{1, \dots, k\}\}$ . In Figure 1 you can see two types of similarity hash functions that based on the random techniques calculations are done. Similarity hash functions are computed by using hash code. Construction of fingerprinting depends on the length of the document  $d$  that must be analyzed at least once, which have the same complexity in all methods, but with each number of fingerprinting, a query  $FD$  document is achieved for document. Therefore, the execution time of retrieving fingerprints is depends on  $K$  and its construction method has two steps: 1. the method that fingerprinting according to the size of the document raises the length of document. 2. The method that  $k$  is independent of  $|d|$  [1].



### III. HEURISTIC ALGORITHMS FOR DETECTING NEAR- DUPLICATE

Today, about 30% of the documents in the web are repeated. According to the same document are more similarities to each other but just in content is the same content. As is clear, duplication and near- duplication of web pages causes problems for search engines, which makes the Users unsatisfied. This requires the creation of efficient algorithms for computing repeated clusters which the goal is to find duplicate clusters. For this purpose, two syntactic and lexical approaches are applied. In syntactic approach only the syntactic structure of documents are examined. This means that the meaning of words contained in the documents is not reviewed, and just the existing of the same words without attention to the meaning.

of them is sufficient for announcing duplication in documents. One of the reasons for increasing documents and similar availability in the Web is easily access to data in web and the lack of semantic method in detection the same availabilities. Also, the decision about the reliability of these documents, when a different version of it with the same content is available, it will be difficult. Every key word in a document and a query user may have different meanings. Therefore, only apparent similarity measurement of documents cannot give the best results to the user query. Most of the current approaches are based on the words semantic properties, such as the relationship between them. Therefore, needed to use semantic in purpose of meaningful comparison in identifying similar documents is feeling [2,13]. Algorithms that are adaptive with intelligent technology, and use heuristic approaches (heuristics) are shown in Figure 2. Among these algorithms:

#### 3.1. SPEX

The basic idea of SPEX operations is that we can show uniquely each sub-segment from a certain piece, and then the whole piece is unique [3].

#### 3.2. I-Match

This algorithm calculates inverse document frequency weighting in order to extract words. Algorithm I-Match, proposes an algorithm based on (multiple random lexicon) that even to improve the reminder a single sign may also be used [3].

#### 3.3. How to Shingle

Shingle way out is one of the old syntactic approaches, that in order to compute the similarity between two documents was proposed by “Broder“. He pointed to this fact none of the criteria for measuring the distance between the strings measurements are discussed, for example, Hamming distance and edit distance cannot consider duplication so well. In Shingle method each document is broken to overlap pieces that are called «Shingle». In this method Shingles don not rely on any linguistic knowledge except converting document into a list of words. Every text in this way is considered as a sequence of symbols. Every symbol can be a character, word, sentence or paragraph of text. There are no restrictions on the choice of symbol or text units, except that the signs or symptoms of text must be countable. Before any operation, it is assumed that each document using a parser program is converted to a canonical form.

The canonical form is the form in which additional information that may not be of interest, such as punctuation, mark and tags of HTML are removed and then the text will consist of a sequence of symbols.

After preprocessing the document, a list of all Shingles and documents that have appeared in it, is prepared. At this stage, documents are converted into binary that the first value is Shingle itself and the second value is document identifier in which it appears. Shingle size selection usually depends on the desired language and is proportionate to the average length of words in that language. The next stage generates a list of all documents that have common Shingle; the number of shared Shingles is also recorded in this stage. In fact the input to this phase is an ordered list < amount of Shingle and document identifier > and the output is a ternary form < the number of common Shingle and the first document identifier and the second document identifier >.

Then the ternary based on the first document identifier are merged and sorted, at last the total number of common Shingles of two documents regarding the measure similarity of the proposed relationships are reviewed. If the similarity is greater than the threshold, two documents are considered as the equivalent and finds it almost duplicated.

Relationships can help to reduce the allocated space has been sampled from a set of Shingle. Also, we can attribute a length1 specific identifier to every Shingle. This mapping is usually done by Rabin fingerprints .The choice of the length 1 is very important because considering the length less than 1 increases the risk of collision and greater than 1 increases the storage space [4, 5].

### **3.4. Shingle Method with large size**

Shingle method with large size, is re-Shingling of Shingles. In fact, in this method at first, the Shingle method outlined in the previous section on applies to any document. Then abstract obtained on a Shingle is sorting and re- arranged method applies on sorting Shingles. In other words, in this way every text is displayed by Shingles with large sizes. In this way, having just one Shingle with large size common is almost sufficient for announcing two documents duplicated. Have a Shingle with equal large size in the two documents is defined as having a common Shingle sequence [6].

In this method in contrast to conventional Shingle methods, there is no need to be collected and counted common Shingles. So, this method for comparing similarities to normal Shingle is simpler and more efficient. But the problem is that this method does not work well for small documents, because it is difficult to prepare Shingle large size of a small document and doesn't have the usual accuracy Shingle. In addition, this method is not able to identify the scope [4]. However, this method before re- Shingle, arranges existing Shingle. Indeed, this is a sampling of available Shingle. It should be noted that both methods have high positive error [6].

### **3.5. Hashed breakpoints**

According to that hashed value is significant breakpoints for searching; large collections of documents can be used. Each word in its hash value (for example, the total number of ASCII in words) is divided into  $n$  parameters [7].

### **3.6. Winnowing**

This algorithm is for selecting strings hash fingerprint for  $k$  gramme. Winnowing selects a document that has little similarity with other documents. If the sub-strings have similarity, a threshold guarantee for similar items is considered. Winnowing procedure at each step selects the minimum hash value [8].

### **3.7. Random (random design)**

This algorithm is used the cosine similarity relationship between the array of words. This algorithm produces an array of binary with  $m$  bits to represent documents. The way it works is that each unique word in the target document is written to a random  $m$ -dimensional array, where each element randomly is chosen from  $[-1,1]$ . Then all of the generated random arrays of words in the document of the previous stage are added together. The  $m$ -dimensional array is produced from results that are accumulated before. Now, with each array element if its value is positive, one element, and if the value is negative, zero element is located instead. Random sampling needs  $\Omega(n)$  to provide exact estimates. When the size of the data set increases, the accuracy of the random projection method algorithm greatly reduced. This is because, random sampling of pairs showing almost a zero proliferative. It is expected because when the size of the data set will increase, the size of the random sample does not alter and the probability that pairs of proliferative appear in random samples is reduced. While dataset is large and sample size is small, it is very likely that almost any pair of samples is not protected. These conditions, even when the data set contains less pairs that are almost duplicated, will be worsening [9]. This is the point where this method in comparison with the Shingle often has less positive error rate. Ignoring other words, the number of occurrences of words and their weights is the problem of this approach [10].

### **3.8. The independent permutation procedure (min-wise)**

As noted above, calculating the similarity between two documents which the size of Shingle set is large is difficult. Min-wise independent permutation algorithm to solve this problem was proposed by Broder. In this algorithm, the degree of similarity between the hashed values is calculated with Jaccard relationship. The procedure of this algorithm is that any set of Shingle is mapped into  $m$ -dimensional array that  $m$  is much smaller than the actual number of symbols in a document. In this process,  $m$  different hashed function with the names of  $h_1, h_m$ , are produced and will be applied to all Shingle. If the final document displays with  $S(n) = \{s_1, s_2, s_n\}$ ,  $J_{th}$  element of this set can show the lowest level hash of pervious stage [11].

### **3.9. Locality Sensitive Hashing Algorithm (LSH)**

In method LSH some hash function is used to determine similar documents. The procedure is that first hash functions are classified to  $k$  Triad bands in other words, each band consists of  $k$  hash functions. All hash functions applied to the input document and the result is stored in each respective band. The hash function is specified for each band and for each pair of documents which are identical to each other. In LSH hash functions in order to combine the results of conditions two approaches are introduced which are known as AND-OR and OR-AND. The method of hash functions into an AND-OR bond with each other AND the results are the output of the AND for each band are the documentation of those couples that like all the bands are known as hash functions. Then the results of different bands are OR together. As a result, the output is each pair that at least by a band of almost duplicated is known [12]. Since the similarity search is an important research issue which is different in application programs for example, media companies such as broadcasters and newspapers are

constantly your pictures and videos uploaded to a repository of multimedia the issue of copyright is one of its main concerns. If the near-duplicated versions are retrieved and reported to the users, the pirated versions are diagnosed quickly. So if new documents are illegal the user must terminate the process of uploaded. Although much research has been done in the context of similarity search, it is still a challenge and to accelerate the search process, in view of N-dimensional similarity space is needed. However, plenty of LSH from space to achieve fast query response is needed. Sim pair LSH is a new approach to speed up the basic LSH method that uses the same parts. Sim pair LSH is better way than LSH because it requires less memory cost [14].

Algorithm	Runtime		Chunk length	Finger-print size	Chunk index size
	Construction	Retrieval			
rare chunks	$O( d )$	$O( d )$	$n$	$O( d )$	$O( d  \cdot  D )$
SPEX ( $0 < r \ll 1$ )	$O( d )$	$O(r \cdot  d )$	$n$	$O(r \cdot  d )$	$O(r \cdot  d  \cdot  D )$
I-Match	$O( d )$	$O(k)$	$ d $	$O(k)$	$O(k \cdot  D )$
shingling	$O( d )$	$O(k)$	$n$	$O(k)$	$O(k \cdot  D )$
prefix anchor	$O( d )$	$O( d )$	$n$	$O( d )$	$O( d  \cdot  D )$
hashed breakpoints	$O( d )$	$O( d )$	$E( c ) = n$	$O( d )$	$O( d  \cdot  D )$
winning	$O( d )$	$O( d )$	$n$	$O( d )$	$O( d  \cdot  D )$
random	$O( d )$	$O(k)$	$n$	$O(k)$	$O( d  \cdot  D )$
one of sliding window	$O( d )$	$O( d )$	$n$	$O( d )$	$O( d  \cdot  D )$
super- / megashingling	$O( d )$	$O(k)$	$n$	$O(k)$	$O(k \cdot  D )$
fuzzy-fingerprinting	$O( d )$	$O(k)$	$ d $	$O(k)$	$O(k \cdot  D )$
locality-sensitive hashing	$O( d )$	$O(k)$	$ d $	$O(k)$	$O(k \cdot  D )$

Figure 2: (complexity of NDD algorithms) [1].

#### IV. ASSESSMENT

Wikipedia by evaluating sets concludes when evaluating of near-duplicate detection methods face the problem for choosing, in standard companies such as TREC or Reuters resemblance is reduced exponentially. It means, documents from the high percentage of very low at similar low percentage of intervals with high similarity are changed. As Figure 3 that shows this feature is shown in Reuters. Figure 3 shows the great size of Wikipedia. The similarity distribution of Reuters and Wikipedia are in conflict with each other.

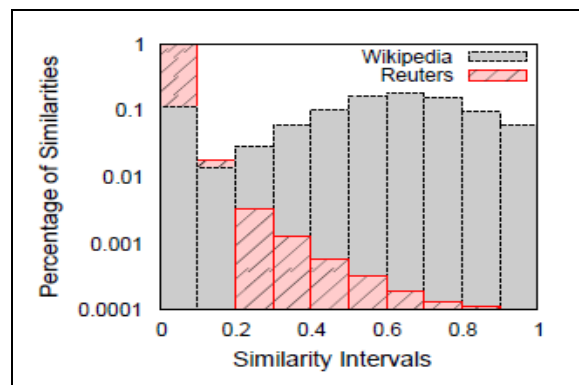
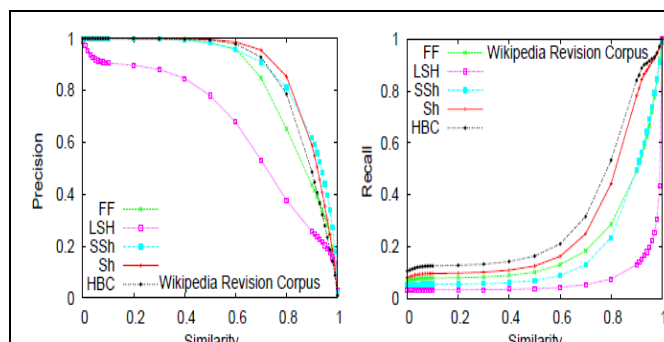


Figure 3: Diagram of the evaluation Wikipedia and Reuters [1].

Figure 4 shows too much similarity universality for fuzzy fingerprint this technique can be used to authenticate and identify that is sensitive to the location of hash and has a significant amount of breakpoint [1].



**Figure 4: Similarity between Documents [1].**

Fingerprinting algorithms are algorithms that are used to detect near-duplication. Wikipedia articles that are regularly listed at the beginning of the address, analyze fingerprinting algorithms of 7 million pairs of documents. And strategies like the first version of each article as a document puts dq search, and compared with other versions it is the first paper that to be replaced. Also near-duplication improves the reliability and accuracy of the assessment (Fig. 4). The results show that pairs of documents that are similar, put into a certain threshold and each set is identified according to the near - duplication. And finally comparison is done by the fingerprint algorithm then the value of integrity using standard methods is computed. The aim is to reduce the ambiguity of user interface. Of course by near-duplication grouping and clustering can hide contents into the cluster. Near-duplication detection is usually done by the search engines on the Web to confirm the approximate fingerprinting, a set of fingerprints that is generated for each document and is used to detect similar documents [1].

The idea of using fingerprints is for identifying duplicates like cryptographic hash functions that are secure against harmful attacks .These functions serve as a high-quality fingerprinting functions. Cryptographic hash algorithms, including MD5 and SHA1 are widely used in the file system. These algorithms are used for data accuracy, and any changes are identified, also approximate fingerprinting detects similar files in a large file system. Fingerprinting normally is used to avoid comparison and transfer great data. In order to realize that remote file browser or proxy server has been modified or not, by fingerprinting and comparison with the previous version, this goal is reached. In other words, virtual fingerprinting is unable to identify a file.

## V. FUTURE WORK

Hash method is a new design for almost the same documents. According to data mining techniques that today are used, times of query is significantly improved, using a combination of heuristic methods and techniques including fingerprinting data encryption technique mining the number of comparisons can be minimized.

## VI. RESULTS

Fingerprinting and Secrecy algorithms in the relational data base are used in order to protect documents against copyright, and because the reliable identify to near-duplicate documents must be existed so, run time is high. Also near-duplicated of the web pages on the search engines have a combination of problems including user dissatisfaction. For this reason, heuristic algorithms such as: SPEX, Shingle off, Shingle large size, hashed breakpoints, Winnowing, Random, min-wise are used. By comparison that is done between Wikipedia and Reuters; the similarity between documents exponentially is decreased. It was shown that the distribution of similarities Reuters and Wikipedia are in conflict with each other. Cryptographic hash algorithms for data verification are used to identify any changes. Also approximate fingerprinting detects similar files in a large file system. Thus, the results show that a set of documents that are similar pairs are placed at a certain threshold, and each set according to near-duplication is identified. And finally comparison is done by the fingerprint algorithm then the value is calculated using the standard method.

## REFERENCES

- [1] Potthast, M., and Stein, B.: 'New Issues in Near-duplicate Detection', 2008, pp. 601-609
- [2] Ignatov, D., J'anos-Rancz, K., and Kuznetsov, S.: 'Towards a framework for near-duplicate detection in document collections', 2009, pp. 215-233
- [3] Pamulaparty, L., Rao, D.M.S., and Rao, D.C.V.G.: 'A Survey on Near Duplicate Web Pages for Web Crawling', 2013, 2 (9), pp. 1-6
- [4] Broder, A.Z.: 'On the resemblance and containment of documents', 1997, pp. 1-9
- [5] Giancarlo, R., and Sanko, D.: 'Identifying and Filtering Near-Duplicate Documents', 2000, pp. 1-10
- [6] Chowdhury, A.K.A.: 'Lexicon randomization for near-duplicate detection with I-Match', 2008, pp. 255-276
- [7] Finkel, R.A., Zaslavsky, A., Monostori, K., and Schmidt, H.: 'Signature extraction for overlap detection in documents', 2001, 4, pp. 1-6
- [8] Schleimer, S., Wilkerson, D.S., and Aiken, A.: 'Winnowing: Local Algorithms for Document Fingerprinting', 2003, pp. 76-85
- [9] Deng, F., and Rafiei, D.: 'Estimating the Number of NearDuplicate Document Pairs for Massive Data Sets using Small Space', 2007, pp. 1-10
- [10] Fan, J., and Huang, T.: 'A fusion of algorithms in near duplicate document detection', 2012, pp. 234-242
- [11] Broder, A.Z., Charikar, M., Frieze, A.M., and Mitzenmacher, M.: 'Min-Wise Independent Permutations', 1998, pp. 1-36
- [12] Gionis, A., Indyk, P., and Motwani, R.: 'Similarity Search in High Dimensions via Hashing', 1999, pp. 1-12
- [13] Alsulami, B.S., Abulkhair, M.F., and Eassa, F.E.: 'Near Duplicate Document Detection Survey', *Computer Science & Communication Networks*, 2012, 2(2), pp. 147-151
- [14] Fisichella, M., Deng, F., and Nejdl, W.: 'Efficient Incremental Near Duplicate Detection Based on Locality Sensitive Hashing', 2010, pp. 152-166

# Mathematical Modeling of Class B Amplifier Using Natural and Regular Sampled Pwm Modulation

<sup>1</sup>, N. V. Shiwardkar, <sup>2</sup>, K. G. Rewatkar\*

Department of Electronics, Dr. Ambedkar College, Deeksha Bhoomi, Nagpur, India

\*Department of Physics, Dr. Ambedkar College, Deeksha Bhoomi, Nagpur India

## ABSTRACT

Class-D amplifiers operate by converting an audio input signal into a high-frequency square wave output, whose lower-frequency components can accurately reproduce the input. Their high power efficiency and potential for low distortion makes them suitable for use in a wide variety of electronic devices. By calculating the outputs from a classical class-D design implementing different sampling schemes we demonstrate that a more advance method, over the double Fourier series method, which is the traditional technique employed for this analysis. This paper shows that when natural sampling is used the input signal is reproduced exactly in the low-frequency part of the output, with no distortion. Although this is a known result, our calculations present the method and notation that develops the classical class-D design is prone to noise, and therefore negative feedback is often included in the circuit. Subsequently we incorporate the Fourier transform/Poisson Re-summation method into a formulised and analysis of a feedback amplifier. Using perturbation expansions we derive the audio-frequency part of the output, demonstrating that negative feedback introduces undesirable distortion.

**KEYWORDS:** class D, natural sampling, regular sampling, Fourier analysis method re-summation

## I. INTRODUCTION

Amplifiers are used increasingly in our every day appliances. In many of the applications efficiency is highly desirable to reduce power consumption. This is important not only from an environmental and cost perspective, but also to maximize battery life on portable devices. Traditional audio amplifiers can achieve efficiencies only in the region of 65-70%, whereas class-D amplifiers can achieve over 90% efficiency [1, 2]. Their high power efficiency, and dissipation less energy is dissipate, there is no need for a large heat sink, means they are suited for use in very small devices, or those where a long battery life is essential, e.g. mobiles, laptops, hearing aids and MP3 players, as well as home sound systems. The key feature of class-D amplifiers that provides such high efficiency is that they are switching amplifiers. This means that their output is a high-frequency square wave that alternates between two voltages. While efficiency is desirable, it is also vital that the amplifier output has low distortion. Theoretically a classical class-D amplifier is able to reproduce an input signal with no distortion at all. It has long been known that this is the case if a sinusoidal signal is input [3], and has been shown more recently for a general input signal [4]. Class-D amplifiers have been implemented commercially. Since the transistors were readily available in the early 1990s [5].

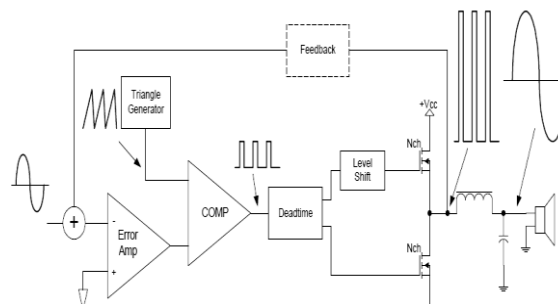


Fig:1 block diagram of class D amplifier

Counter-intuitive though it may seem, the square wave output from class-D amplifiers can reproduce a sound free from distortion, and in a highly efficient manner. A high frequency square wave is the most efficient output, much more efficient than a slowly varying output signal where a lot of energy would be dissipated as heat.

In order to understand how the square wave output from the amplifier provides low distortion, it is important to examine the performance of amplifier is used. Class-D amplifiers are used in the output stage (see figure 1). A pre-amplifier first increases the amplitude of the low-amplitude analogue audio signal. The signal, now at the required amplitude for playback, then passes through a class-D amplifier, which converts the signal into a more efficient form (a square wave) for playback. The square wave then passes through a filter and a loudspeaker, which plays the final output signal in its amplified form. Therefore, rather than to increase amplitude, the aim of a class-D amplifier is to convert the input signal into a square wave that represents the input signal. To do this, a class-D amplifier creates a square wave whose width varies according to the input signal, via a process called pulse width modulation (PWM)[3,7]. The way this process is carried out is important because after filtering, the output should ideally equal the signal input to the class-D amplifier. When PWM is used, a relatively low-frequency input signal is compared with a carrier wave of much higher frequency to create a high-frequency square wave that switches between voltages  $+V$  and  $-V$ . The widths of the pulses in this resulting

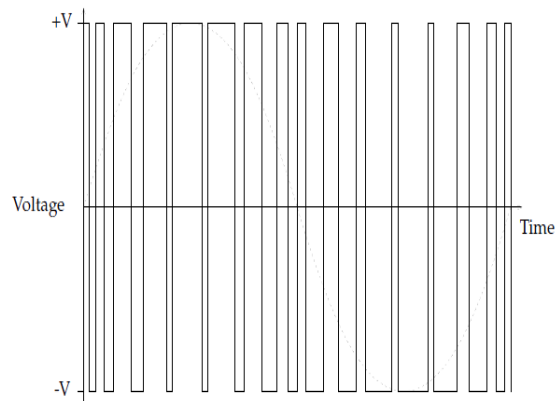


Fig: 2 PWM cycle swing between  $+v$  to  $-v$

Thus the pulse width modulated square wave is composed of low-frequency components related to the input signal, and high-frequency components related to the carrier wave. The square wave is then passed through a filter where the high-frequency components are attenuated, while the low-frequency components are allowed to pass through relatively unchanged. These low-frequency components constitute the final output, which is as close to the original input signal as possible. The duty cycle of this square wave is defined as the ratio between the length of time the wave is at  $+V$  and the period of the carrier wave. An ideal square wave, which is at  $+V$  for half of the period and  $-V$  for the remaining half, has a 50% duty cycle. Therefore we see that when PWM is used it is the duty cycle of the square wave output which varies according to the input signal. The typical frequency ranges are 80-250 kHz for the carrier wave [1], and 20Hz-20 kHz for the input signal [6]

### Experimentation

The classical class- D Amplifier design is known to reproduce the input signal exactly in the low-frequency part of the output with no distortion [1 7]. However, this simple design is susceptible to noise, for example due to non-ideal components, or variation in the carrier wave [8]. For this reason, negative feedback is often implemented in class-D designs. Negative feedback allows the output to be “fed back” into the circuit, in order to achieve a final output that is closer to the input signal. While negative feedback reduces noise in the circuit, it does however introduce distortion [7] In order to investigate class-D amplifiers, it is therefore essential to analyse the square wave output produced by PWM. However, it is not straightforward to achieve this because, even for a periodic or quasi-periodic input, the output is only quasi-periodic. we calculate the pulse width modulated output from a classical class-D amplifier, and show that the input signal can be reproduced exactly in the theoretical output[8,9]. The method of calculating output, in order to highlight the advantages of the second. The first method we implement is the double Fourier series method [3]. Natural sampling and regular sampling (which is sometimes referred to as uniform sampling) are two commonly used methods of sampling[11] which is the most commonly used. However, the approach is unnecessarily complex and it is circuitous, though possible, to extend the method to more advanced modulation schemes [12, 13]. which is simpler and can be adapted easily to investigate other modulation schemes.

II. ANALYSIS

Natural sampling and regular sampling (which is sometimes referred to as uniform sampling) are two commonly used methods of sampling. We now show how these methods can be implemented to create a pulse width modulated square wave output,  $g^*(t^*)$ , that alternates between  $+V$  and  $-V$ . The switching times of  $g^*(t^*)$  are determined by the intersection of the input signal  $s^*(t^*)$  with a high-frequency carrier wave  $v^*(t^*)$  of period  $T$  and (angular) frequency  $\omega^*c = 2\pi/T$  the carrier wave can be defined in different ways, according to the type of modulation required, as we shall see below. When natural sampling is used, the switching occurs when  $s^*(t^*) + v^*(t^*) = 0$ . When regular sampling is used, the input signal is sampled at a fixed time in each carrier wave period, and the switching occurs -when this sample equal-minus the carrier wave. For example, if the input signal is sampled at the beginning of each carrier wave period, when  $t^* = nT$ , the switching occurs at a time  $t^*$  later in that carrier wave period when  $s^*(nT) + v^*(t^*) = 0$ .

It is possible to use either single-edge or double-edge modulation. When single- edge modulation is used, only one edge of the square wave output is determined by the input signal, the other edge occurs at a fixed time. The leading edge of the square wave is defined as the one that switches from  $-V$  to  $+V$ , and the trailing edge is defined as the one that switches from  $+V$  to  $-V$ .

If leading-edge modulation used, the leading edge is determined by the input signal and the trailing edge remains fixed. If trailing-edge modulation is used, then its edge is determined by the input signal and it remains fixed[14]. For single-edge modulation the carrier wave is a sawtooth wave with period  $T$ , where for leading-edge modulation it is defined to be

$$v^*(t^*) = -V + \frac{2V}{T}(t^* - nT) \text{ for } nT < t^* < (n+1)T,$$

-----(1)

$$v^*(t^*) = V - \frac{2V}{T}(t^* - nT) \text{ for } nT < t^* < (n+1)T,$$

-----(2)

Equation for leading edge and trailing edge input signal.

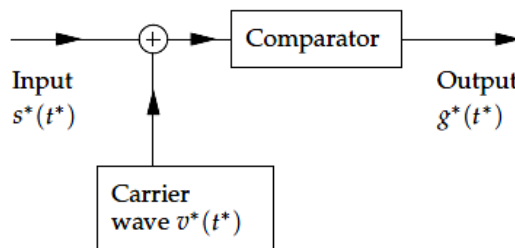


Fig.3: modeling for traling edge or leading edge class d amplifire

4.1 Equation basis analysis

A classical model of class-D amplifier, as depicted in figure 3. The input signal  $s^*(t^*)$  is first added to a carrier wave  $v^*(t^*)$ . The resulting voltage,  $s^*(t^*) + v^*(t^*)$ , is fed into a comparator that produces a square wave output,  $g^*(t^*)$ , defined by equation (3) as

$$g^*(t^*) = \begin{cases} -V & \text{for } s^*(t^*) + v^*(t^*) < 0 \\ +V & \text{for } s^*(t^*) + v^*(t^*) > 0. \end{cases}$$

-----(3)

Where appropriate you should overwrite the different fields When natural sampling is used, the switching occurs when  $s^*(t^*) + v^*(t^*) = 0$ . When regular sampling is used, the input signal is sampled at a fixed time in each carrier wave period, and the switching occurs when this sample equals minus the carrier wave. For example, if the input signal is sampled at the beginning of each carrier wave period, when  $t^* = nT$ , the switching occurs at a time  $t^*$  later in that carrier wave period when  $s^*(nT) + v^*(t^*) = 0$ .



Natural and regular sampling has been investigated extensively, and it is well documented that natural sampling produces less distortion than regular sampling. This has been shown for a general input signal [4], and has also been verified for particular input signals, [15–19]. investigate this classical class-D amplifier design for a sinusoidal input signal, defined to be

$$S^*(t^*) = S_0 V \sin \omega^* a t^* \text{-----(4)}$$

The carrier wave is therefore a sawtooth wave defined by equation(4) and figure 3, and we apply natural and regular sampling as depicted in figure 3. The square wave  $g^*(t^*)$  switches from  $-V$  to  $+V$  at times  $t^* = nT + \beta^*n$  and from  $+V$  to  $-V$  at times  $t^* = nT$ , and therefore we may write

$$g^*(t^*) = \begin{cases} -V & \text{for } nT < t^* < nT + \beta_n^* \\ +V & \text{for } nT + \beta_n^* < t^* < (n + 1)T. \end{cases} \text{-----(5)}$$

When natural sampling is used, the leading-edge switching occurs when  $s(t) + v(t) = 0$ . Thus for natural sampling we have

$$\beta_n = \frac{1}{2}(1 - s(n + \beta_n)). \text{-----(6)}$$

When regular sampling is used, the input signal is sampled at the beginning of each carrier wave period,  $t = n$ . The leading-edge switching occurs when  $s(n) + v(t) = 0$ . Thus for regular sampling we have

$$\beta_n = \frac{1}{2}(1 - s(n)). \text{-----(7)}$$

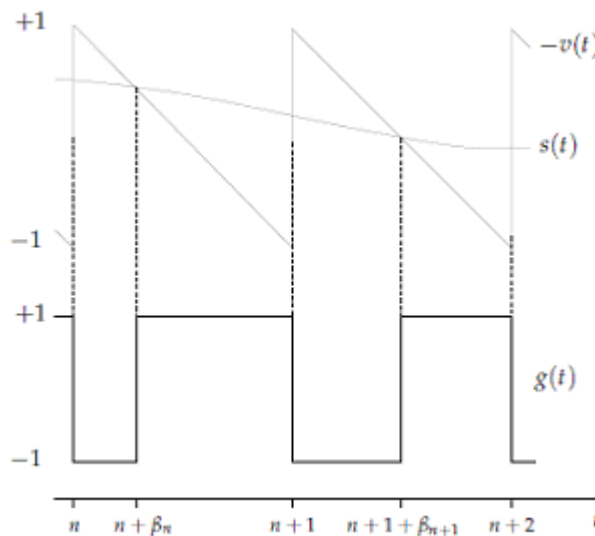


Fig 4: wave shaped and applied technology

An effective way to analyse the output resulting from particular sampling scheme is to plot the spectrum, i.e. plot the magnitude of the amplitude of each coefficient  $G_{mn}$  in the output against its frequency  $\omega$ . By plotting and comparing the spectra of different sampling schemes we can see clearly what the components of each output are, and their magnitudes, and so determine which sampling scheme produces the output with lower distortion. The spectrum for regular sampling is plotted in figure 4 next to the spectrum for natural sampling. Note that to plot both spectra on the same graph we have shifted the spectrum for regular sampling to the right by 0.05, so that, for example, the black peak that appears at  $\omega = 0.3$  is actually the peak that corresponds to  $\omega = 0.25$ . It is seen from figure 4 that the only component in the low-frequency part of the natural sampling spectrum is exactly the input signal, whereas for regular sampling the input signal harmonics of the input signal appear in the low-frequency part of the output. These harmonics can be seen more clearly in figure 4 where we plot only the low-frequency part of the output. Thus comparing the low-frequency parts of the spectra for natural and regular sampling, it is obvious that the output resulting from regular sampling contains much more distortion

than that from natural sampling, which contains no distortion. Outside the low-frequency part of the spectrum, the outputs from both sampling schemes comprise peaks at multiples of the carrier wave frequency as well as lower amplitude peaks (called side-bands) concentrated around multiples of the carrier wave frequency. Note that we have chosen to plot the spectra up to  $\omega = 16$  in figure 5 merely so that the low-frequency part of the spectra, as well as the peaks at  $\omega = 2\pi$  and  $\omega = 4\pi$  (i.e. at the carrier wave frequency and at twice the carrier wave frequency) and their corresponding sidebands, can be seen clearly. In addition there are peaks at, and sidebands around, all larger multiples of the carrier wave frequency, as can be determined from the natural and regular sampling output formulae. There are minor differences in the amplitudes of these peaks outside the low-frequency part of the spectrum for regular sampling compared with natural sampling, but these are irrelevant as they will be attenuated by a low-pass filter.

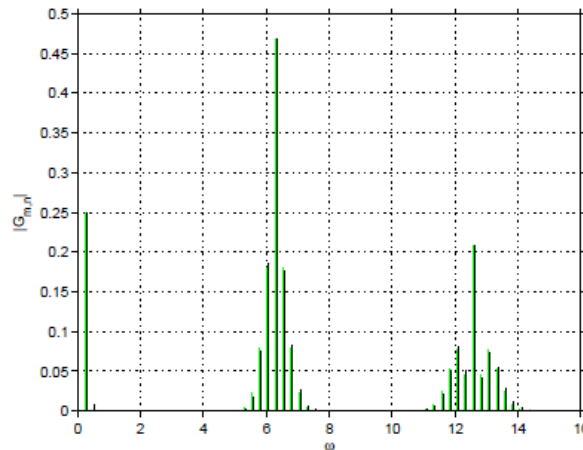


Fig. 5 spectrum of low frequency PWM

### III. DISCUSSION

The response of the design is determined the outputs from a classical class-D amplifier when a sinusoidal signal is input and leading-edge natural or regular sampling PWM is used to create the square wave output. For natural sampling, the input is reproduced exactly in the output, and there are no other terms in the low- frequency part of the output, and so the input signal can be reproduced with no distortion. For regular sampling, the input signal is reproduced with distortion and harmonics of the input signal appear in the low-frequency part of the output, and so the amplifier output is a distorted version of the input signal. Although natural sampling provides better distortion performance than regular sampling, it is only suited to some applications. The equations for the natural sampling switching times are implicit, and so natural sampling is often used in analogue applications, but is difficult to implement digitally [16].

The equations for the regular sampling switching times are explicit and so this sampling scheme is commonly used in digital applications. This motivates us to investigate sampling schemes that aim to provide low distortion, like natural sampling, whilst being simple to use in digital applications, like regular sampling. The calculated outputs for the two sampling schemes first using the commonly used double Fourier series method, and then repeated the calculations using the Fourier transform/Poisson Re-summation method, in order to illustrate the advantages of the latter method. When the comparison of the two methods for natural sampling, it is easy to see that the latter method is simpler and quicker to implement. Not needing to introduce two separate timescales to the problem and using the Poisson resumption formula shortens the calculation considerably. In addition, it is possible to demonstrate via the Fourier transform/Poisson resumption method for a general input signal that the low-frequency part of the output for natural sampling is exactly the input signal, which is not possible via the double Fourier series method. Both methods require adaptation to examine the output resulting from regular sampling. However, using the double Fourier series method, an additional change of variables is required to solve the problem, making the method unnecessarily complex. The alteration to the Fourier transform/Poisson re-summation method is to take the Fourier transform, which is simple to invert later. This change ensures that the equations for the switching times are used in their explicit form. Each method requires separate consideration of particular frequency components, though using the Fourier transform/Poisson re-summation method this can be done quickly, especially in the natural sampling case. The Fourier transform/Poisson re-summation method has considerable advantages over the double Fourier series method. It is shorter and simpler to use, as well as being more easily adaptable to different sampling schemes.

It enables easy comparisons between existing modulation and sampling techniques, as well as mathematical analysis of new or complex strategies that so far have not been tackled.

#### IV. CONCLUSION

With the detailed result analysis considered, the PWM process and investigated the methods by which pulse width modulated square waves can be analyzed. The two different approaches to analyzing the pulse width modulated output created by a classical class-D amplifier. The double Fourier series method, which is the conventional technique, was shown to be unnecessarily complex. We demonstrated that the Fourier transform/Poisson resumption method is simpler and quicker to implement, as well as being easier to adapt to different sampling schemes, in the analysis of the classical class-D amplifier, it determined that, if natural sampling is used, the low-frequency part of the output is exactly the input signal. However, the classical design is susceptible to noise, and negative feedback is often included in the circuit to counter this problem. We devote the rest of the investigation of negative feedback designs, and incorporate the Fourier transform/Poisson resumption method into each model.

#### REFERENCES

- [1] C. Pascual, Z. Song, P. T. Krein, D. V. Sarwate, P. Midya, and W. J. Roeckner. High fidelity PWM inverter for digital audio amplification: spectral analysis, real-time DSP implementation, and results. *IEEE Transactions on Power Electronics*, 18(1): 473–485, 2003.
- [2] M. Berkhout and L. Dooper. Class-D audio amplifiers immobile applications. *IEEE Transactions on Circuits and Systems-I: Regular Papers*, 7(5):992–1002, 2010.
- [3] H. S. Black. *Modulation Theory*. Van Nostrand, New York, 1953.
- [4] Z. Song and D. V. Sarwate. The frequency spectrum of pulse width modulated signals. *Signal Processing*, 83:2227–2258, 2003.
- [5] B. Putzeys. Digital audio's final frontier. *IEEE Spectrum*, 40(3):34–41, 2003.
- [6] D. J. Benson. *Music: A Mathematical Offering*. Cambridge University Press, Cambridge, 2006.
- [7] S. M. Cox and B. H. Candy. Class-D audio amplifiers with negative feedback. *SIAM Journal on Applied Mathematics*, 66(2):468–488, 2005.
- [8] M. T. Tan, J. S. Chang, H. C. Chua, and B. H. Gwee. An investigation into the parameters affecting total harmonic distortion in low-voltage low-power class-D amplifiers. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 50(13):1304–1315, 2003.
- [9] W. R. Bennett. New results in the calculation of modulation products. *Bell System Technical Journal*, 12:228–243, 1933.
- [10] G. Fedele and D. Frascino. Spectral analysis of a class of DC-AC PWM inverters by Kapteyn series. *IEEE Transactions on Power Electronics*, 25(4):839–849, 2010.
- [11] S.M. Cox and S. C. Creagh. Voltage and current spectra for matrix power converters. *SIAM Journal on Applied Mathematics*, 69(5):1415–1437, 2009.
- [12] J. T. Boys and P. G. Handley. Harmonic analysis of space vector modulated PWM waveforms. *IEE Proceedings, Part B*, 137(4):197–204, 1990.
- [13] D. G. Holmes. A general analytical method for determining the theoretical harmonic components of carrier based PWM strategies. In *Conf. Rec. IEEE-IAS Annual Meeting*, pages 1207–1214, 1998.
- [14] H. Li, B. H. Gwee, and J. S. Chang. A digital class D amplifier design embodying a novel sampling process and pulse generator. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pages 826–829, 2001.
- [15] R. A. Guinee and C. Lyden. A novel Fourier series time function for modeling and simulation of PWM. *IEEE Transactions on Circuits and Systems-I: Regular Papers*, 52 (11):2427–2435, 2005.
- [16] D. G. Holmes and T. A. Lipo. *Pulse Width Modulation for Power Converters: Principles and Practice*. IEEE Press, Piscataway, NJ, 2003.

## Neighborhood Triple Connected Domination Number of a Graph

G. Mahadevan<sup>1</sup> N. Ramesh<sup>2</sup> C. Sivagnanam<sup>3</sup> Selvam Avadayappan<sup>4</sup>  
A. Ahila<sup>5</sup> T.Subramanian<sup>6</sup>

<sup>1</sup>Dept. of Mathematics, Gandhigram Rural Institute – Deemed University,  
Gandhigram – 624 302

<sup>2</sup>Udaya School of Engineering, Udaya Nagar, Vellamodi,  
Kanyakumari – 629 204

<sup>3</sup>Dept. of General Requirements, College of Applied Sciences – lbri,  
Sultanate of Oman. Postal Code - 516

<sup>4</sup>Dept. of Mathematics, VHNSN College, Virudhunagar – 626 001

<sup>5</sup>Dept. of Mathematics, Science and Humanities Department,  
S. Veerasamy Chettiar College of Engineering and Technology, Puliangudi,  
Tirunelveli – 627 855

<sup>6</sup>Dept. of Mathematics, Anna University : Tirunelveli Region,  
Tirunelveli – 627 007

### ABSTRACT:

In this paper we introduce new domination parameter with real life application called neighborhood triple connected domination number of a graph. A subset  $S$  of  $V$  of a nontrivial graph  $G$  is said to be a neighborhood triple connected dominating set, if  $S$  is a dominating set and the induced subgraph  $\langle N(S) \rangle$  is a triple connected. The minimum cardinality taken over all neighborhood triple connected dominating sets is called the neighborhood triple connected domination number and is denoted by  $\gamma_{ntc}$ . We investigate this number for some standard graphs and find the lower and upper bounds of this number. We also investigate its relationship with other graph theoretical parameters.

**KEY WORDS:** Neighborhood triple connected domination number

AMS (2010): 05C69

### I. INTRODUCTION

By a **graph** we mean a finite, simple, connected and undirected graph  $G(V, E)$ , where  $V$  denotes its vertex set and  $E$  its edge set. Unless otherwise stated, the graph  $G$  has  $p$  vertices and  $q$  edges. **Degree** of a vertex  $v$  is denoted by  $d(v)$ , the **maximum degree** of a graph  $G$  is denoted by  $\Delta(G)$ . A graph  $G$  is **connected** if any two vertices of  $G$  are connected by a path. A maximal connected subgraph of a graph  $G$  is called a **component** of  $G$ . The **complement**  $\bar{G}$  of  $G$  is the graph with vertex set  $V$  in which two vertices are adjacent if and only if they are not adjacent in  $G$ . We denote a **cycle** on  $p$  vertices by  $C_p$ , a **path** on  $p$  vertices by  $P_p$ , and a **complete graph** on  $p$  vertices by  $K_p$ . A **wheel graph**  $W_n$  of order  $n$ , sometimes simply called an  $n$ -wheel, is a graph that contains a cycle of order  $n-1$ , and for which every vertex in the cycle is connected to one other vertex. A **tree** is a connected acyclic graph. The complete bipartite graph with partitions  $|V_1|=m$  and  $|V_2|=n$ , is denoted by  $K_{m,n}$ . A **star**, denoted by  $K_{1,p-1}$  is a tree with one root vertex and  $p-1$  pendant vertices. A **bistar**, denoted by  $B(m,n)$  is the graph obtained by joining the root vertices of the stars  $K_{1,m}$  and  $K_{1,n}$ . The **friendship graph**, denoted by  $F_n$  can be constructed by identifying  $n$  copies of the cycle  $C_3$  at a common vertex. A **helm graph**, denoted by  $H_n$  is a graph obtained from the wheel  $W_n$  by joining a pendant vertex to each vertex in the outer cycle of  $W_n$  by means of an edge.

The **cartesian graph product**  $G = G_1 \square G_2$ , sometimes simply called the graph product of graphs  $G_1$  and  $G_2$  with disjoint point sets  $V_1$  and  $V_2$  and edge sets  $X_1$  and  $X_2$  is the graph with point set  $V_1 \times V_2$  and  $u = (u_1, u_2)$  adjacent with  $v = (v_1, v_2)$  whenever  $[u_1 = v_1 \text{ and } u_2 \text{ adj } v_2]$  or  $[u_2 = v_2 \text{ and } u_1 \text{ adj } v_1]$ . The  **$m$ -book graph**  $B_m$  is defined as the graph Cartesian product  $S_{m+1} \times P_2$ , where  $S_m$  is a star graph and  $P_2$  is the path graph on two nodes.

The **open neighborhood** and **closed neighborhood** of a vertex  $v$  are denoted by  $N(v)$  and  $N[v] = N(v) \cup \{v\}$  respectively. If  $S \subseteq V$ , then  $N(S) = \bigcup_{v \in S} N(v)$  and  $N[S] = N(S) \cup S$ . A **cut – vertex (cut edge)** of a graph  $G$  is a vertex (edge) whose removal increases the number of components. A **vertex cut**, or **separating set** of a connected graph  $G$  is a set of vertices whose removal results in a disconnected graph. The **connectivity** or **vertex connectivity** of a graph  $G$ , denoted by  $\kappa(G)$  (where  $G$  is not complete) is the size of a smallest vertex cut. The **chromatic number** of a graph  $G$ , denoted by  $\chi(G)$  is the smallest number of colours needed to colour all the vertices of a graph  $G$  in which adjacent vertices receive different colour. A **Nordhaus -Gaddum-type** result is a (tight) lower or upper bound on the sum or product of a parameter of a graph and its complement. Terms not defined here are used in the sense of [15].

A subset  $S$  of  $V$  is called a **dominating set** of  $G$  if every vertex in  $V - S$  is adjacent to at least one vertex in  $S$ . The **domination number**  $\gamma(G)$  of  $G$  is the minimum cardinality taken over all dominating sets in  $G$ .

Many authors have introduced different types of domination parameters by imposing conditions on the dominating set [19]. In [18] Paulraj Joseph et. al., introduced the concept of triple connected graphs. A graph  $G$  is said to be **triple connected** if any three vertices of  $G$  lie on a path. In [1] Mahadevan et. al., introduced the concept of triple connected domination number of a graph. A subset  $S$  of  $V$  of a nontrivial graph  $G$  is said to be an **triple connected dominating set**, if  $S$  is a dominating set and the induced sub graph  $\langle S \rangle$  is triple connected. The minimum cardinality taken over all triple connected dominating sets is called the **triple connected domination number** of  $G$  and is denoted by  $\gamma_{tc}$ . In [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14] G. Mahadevan et. al., introduced complementary triple connected domination number, complementary perfect triple connected domination number, paired triple connected domination number, triple connected two domination number, restrained triple connected domination number, dom strong triple connected domination number, strong triple connected domination number, weak triple connected domination number, triple connected complementary tree domination number of a graph, efficient complementary perfect triple connected domination number of a graph, efficient triple connected domination number of a graph, complementary triple connected clique domination number of a graph, triple connected.com domination number of a graph respectively and investigated new results on them.

In [21], Arumugam. S and Sivagnanam. C introduced the concept of **neighborhood connected domination in** graph.

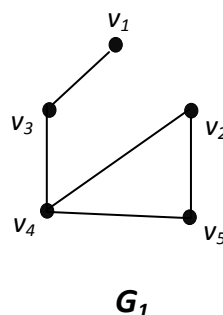
A dominating set  $S$  of a connected graph  $G$  is called a **neighborhood connected dominating set** (ncd-set) if the induced subgraph  $\langle N(S) \rangle$  is connected. The minimum cardinality of a ncd-set of  $G$  is called the **neighborhood connected domination number** of  $G$  and is denoted by  $\gamma_{nc}(G)$ .

Motivated by all the above results and ideas in this paper we introduce new domination parameter called **neighborhood triple connected domination number of a graph**.

## II. NEIGHBORHOOD TRIPLE CONNECTED DOMINATION NUMBER OF A GRAPH

**Definition 2.1** A subset  $S$  of  $V$  of a nontrivial graph  $G$  is said to be a *neighborhood triple connected dominating set*, if  $S$  is a dominating set and the induced subgraph  $\langle N(S) \rangle$  is triple connected. The minimum cardinality taken over all neighborhood triple connected dominating sets is called the *neighborhood triple connected domination number* of  $G$  and is denoted by  $\gamma_{ntc}(G)$ . Any neighborhood triple connected dominating set with  $\gamma_{ntc}$  vertices is called a  $\gamma_{ntc}$ -set of  $G$ .

**Example 2.2** For the graph  $G_1$  in figure 2.1,  $S = \{v_1, v_2\}$  forms a  $\gamma_{ntc}$ -set of  $G_1$ . Hence  $\gamma_{ntc}(G_1) = 2$ .



**Figure 2.1 : Graph with  $\gamma_{ntc} = 2$ .**

**Real Life Application of Neighborhood Triple Connected Domination Number**

Suppose we are manufacturing a product and need to distribute the products in different major cities and sub cities so that we give dealership to each city and the dealers in that city distribute our products in to the sub cities. The major cities may or may not be connected. If we draw this situation as a graph by considering the major cities and sub cities as vertices and the roadways connecting the cities as edges, the cities denote the dominating set say  $S$  of the constructed graph. If  $\langle N(S) \rangle$  is triple connected in the constructed graph means the customer in the sub cities who needs our product when they did not get our product from the corresponding major cities can get our product either from the other major cities or any one of the other sub cities. And also the minimum cardinality of  $S$  minimizes the total cost. The above situation describes one of the real life application of neighborhood triple connected dominating set and neighborhood triple connected domination number of a graph.

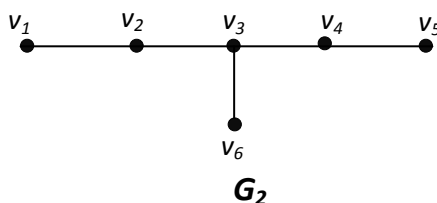
**Observation 2.3** Neighborhood triple connected dominating set ( $\gamma_{ntc}$ -set or ntc set) does not exist for all graphs.

**Example 2.4** For  $K_{1,6}$ , there does not exist any neighborhood triple connected dominating set.

**Remark 2.5** Throughout this paper we consider only connected graphs for which neighborhood triple connected dominating set exists.

**Observation 2.6** The complement of a neighborhood triple connected dominating set  $S$  need not be a neighborhood triple connected dominating set.

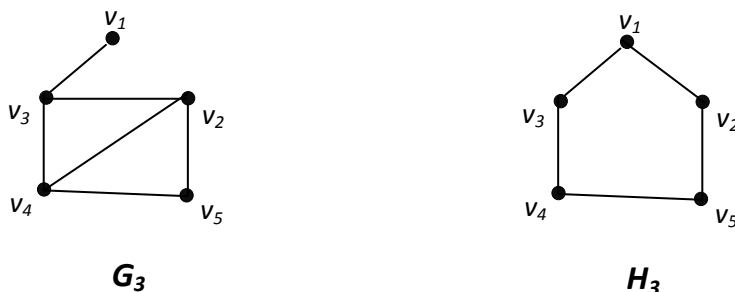
**Example 2.7** For the graph  $G_2$  in the figure 2.2,  $S = \{v_1, v_5, v_6\}$  is a neighborhood triple connected dominating set. But the complement  $V - S = \{v_2, v_3, v_4\}$  is not a neighborhood triple connected dominating set.



**Figure 2.2**

**Observation 2.8** Every neighborhood triple connected dominating set is a dominating set but not conversely.

**Example 2.9** For the graph  $G_3$  in the figure 2.3,  $S = \{v_1, v_2\}$  is a neighborhood triple connected dominating set as well as a dominating set. For the graph  $H_3$  in the figure 2.3,  $S = \{v_3, v_5\}$  is a dominating set but not a neighborhood triple connected dominating set.



**Figure 2.3**

**Exact value for some standard graphs:**

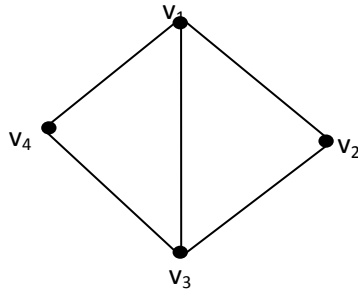
- 1) For any complete graph of order  $p \geq 4$ ,  $\gamma_{ntc}(K_p) = 1$ .
- 2) For any complete bipartite graph of order  $p \geq 4$ ,  $\gamma_{ntc}(K_{m,n}) = 2$  where  $m, n > 1$  and  $m + n = p$ .
- 3) For the wheel graph of order  $p \geq 4$ ,  $\gamma_{ntc}(W_p) = 1$ .

4) For the helm graph  $H_n$  of order  $p \geq 7$ ,  $\gamma_{ntc}(G) = \frac{p-1}{2}$ , where  $2n - 1 = p$ .

**Exact value for some special graphs:**

1) The **diamond graph** is a planar undirected graph with 4 vertices and 5 edges as shown in figure 2.5. It consists of a complete graph  $K_4$  minus one edge.

For any **diamond** graph  $G$  of order 4,  $\gamma_{ntc}(G) = 1$ .

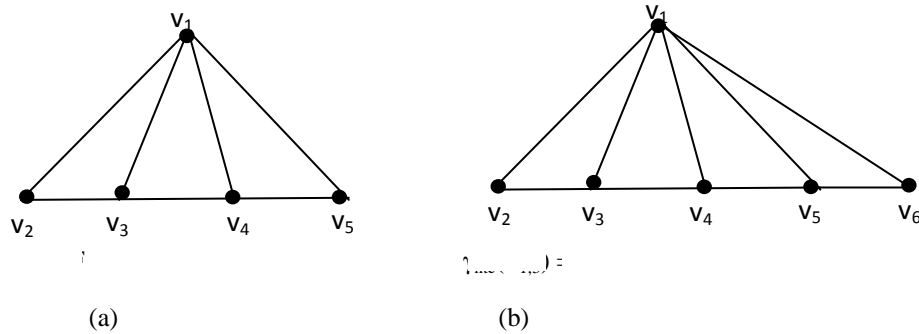


**Figure 2.5**

In figure 2.5 (a)  $S = \{v_1\}$  is a neighborhood triple connected dominating set.

2) A Fan graph  $F_{p,q}$  is defined as the graph join  $\bar{K}_p + P_q$ , where  $\bar{K}_p$  is the empty graph on  $p$  nodes and  $P_q$  is the path graph on  $q$  nodes. The case  $p = 1$  corresponds to the usual fan graphs.

For any **Fan** graph of order  $n \geq 4$ ,  $\gamma_{ntc}(F_{1,n-1}) = 1$ .

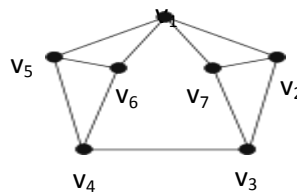


**Figure 2.6**

In figure 2.6 (a)  $S = \{v_1\}$  is a neighborhood triple connected dominating set. And also in figure 2.6 (b)  $S = \{v_1\}$  is a neighborhood triple connected dominating set.

3) The **Moser spindle** (also called the **Mosers' spindle** or **Moser graph**) is an undirected graph with seven vertices and eleven edges as shown in figure 2.7.

For the **Moser spindle** graph  $G$ ,  $\gamma_{ntc}(G) = 2$ .

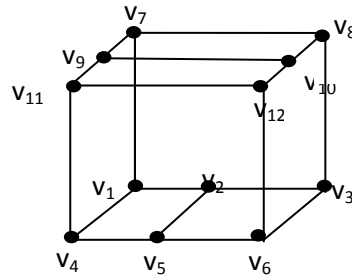


**Figure 2.7**

In figure 2.7,  $S = \{v_3, v_6\}$  is a neighborhood triple connected dominating set.

4) The **Bidiakis cube** is a 3-regular graph with 12 vertices and 18 edges as shown in figure 2.8.

For the **Bidiakis cube** graph  $G$ ,  $\gamma_{ntc}(G) = 4$ .

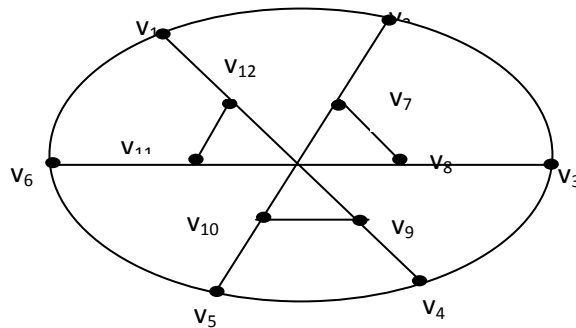


**Figure 2.8**

In figure 2.8,  $S = \{v_1, v_6, v_{10}, v_{12}\}$  is a neighborhood triple connected dominating set.

5) The **Franklin graph** is a 3-regular graph with 12 vertices and 18 edges as shown below in figure 2.9.

For the **Franklin graph**  $G$ ,  $\gamma_{ntc}(G) = 4$ .

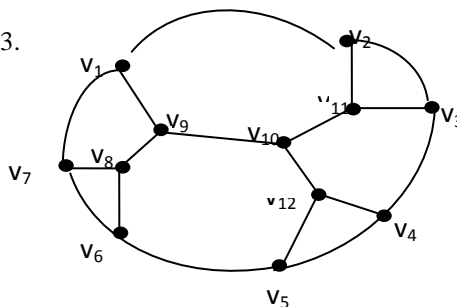


**Figure 2.9**

In figure 2.9,  $S = \{v_4, v_5, v_7, v_{12}\}$  is a neighborhood triple connected dominating set.

6) The **Frucht graph** is a 3-regular graph with 12 vertices, 18 edges, and no nontrivial symmetries as shown below in figure 2.10. It was first described by Robert Frucht in 1939.

For the **Frucht graph**  $G$ ,  $\gamma_{ntc}(G) = 3$ .



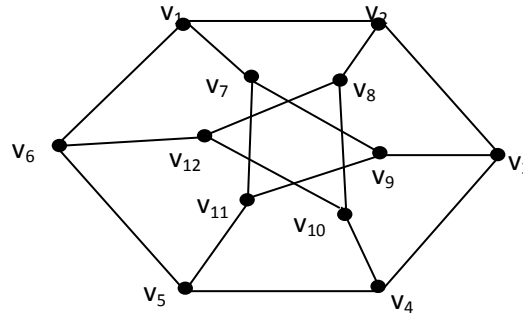
**Figure 2.10**



In figure 2.10,  $S = \{v_2, v_8, v_{12}\}$  is a neighborhood triple connected dominating set.

7) The **Dürer graph** is an undirected cubic graph with 12 vertices and 18 edges as shown below in figure 2.11. It is named after Albrecht Durer.

For the **Dürer graph**  $G$ ,  $\gamma_{ntc}(G) = 4$ .

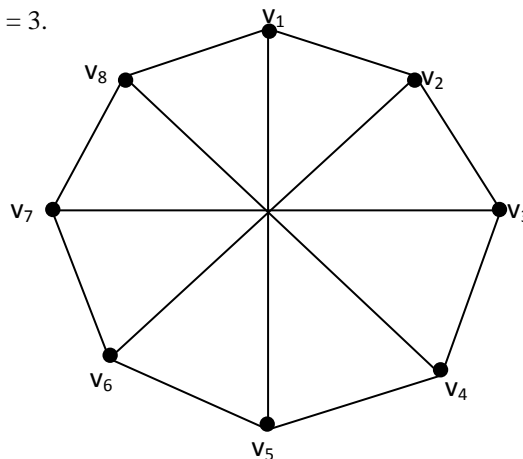


**Figure 2.11**

In figure 2.11,  $S = \{v_5, v_7, v_8, v_9\}$  is a neighborhood triple connected dominating set.

8) The **Wagner graph** is a 3-regular graph with 8 vertices and 12 edges, as shown in figure 2.12, named after Klaus Wagner. It is the 8-vertex Mobius ladder graph. Mobius ladder is a cubic circulant graph with an even number 'n' vertices, formed from an n- cycle by adding edges connecting opposite pairs of vertices in the cycle.

For the **Wagner graph**  $G$ ,  $\gamma_{ntc}(G) = 3$ .

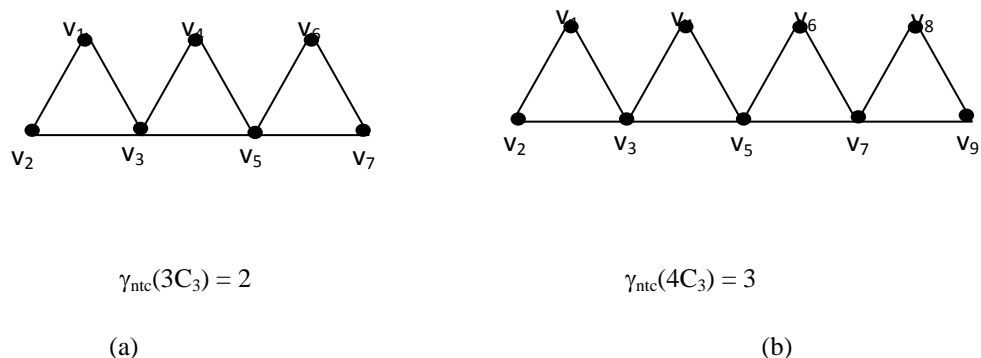


**Figure 2.12**

In figure 2.12,  $S = \{v_2, v_3, v_4\}$  is a neighborhood triple connected dominating set.

9) The **Triangular Snake** graph is obtained from a path  $v_1, v_2, \dots, v_n$  by joining  $v_i$  and  $v_{i+1}$  to a new vertex  $w_i$  for  $i = 1, 2, \dots, n-1$  and denoted by  $mC_3$  (where  $m$  denotes the number of times the cycle  $C_3$ ) snake as shown in figure 2.13.

For the **Triangular Snake**  $G$ ,  $\gamma_{ntc}(G) = \left\lceil \frac{2m}{3} \right\rceil$ .

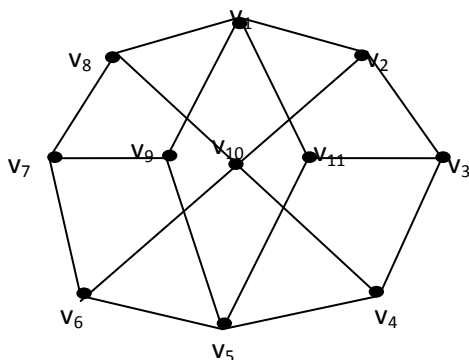


**Figure 2.13**

In figure 2.13 (a)  $S = \{v_3, v_5\}$  is a neighborhood triple connected dominating set. In figure 2.13 (b)  $S = \{v_3, v_5, v_7\}$  is a neighborhood triple connected dominating set.

10) The **Herschel graph** is a bipartite undirected graph with 11 vertices and 18 edges as shown in figure 2.14, the smallest non Hamiltonian polyhedral graph. It is named after British astronomer Alexander Stewart Herschel.

For the **Herschel graph**  $G$ ,  $\gamma_{ntc}(G) = 3$ .

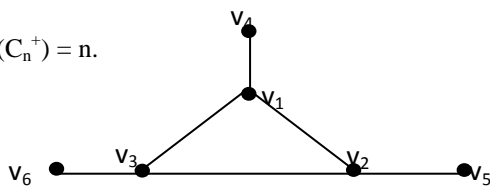


**Figure 2.14**

In figure 2.14,  $S = \{v_9, v_{10}, v_{11}\}$  is a neighborhood triple connected dominating set.

11) Any cycle with a pendant edge attached at each vertex as shown in figure 2.15 is called **Crown graph** and is denoted by  $C_n^+$ .

For the **Crown graph**,  $\gamma_{ntc}(C_n^+) = n$ .



**Figure 2.15**

In figure 2.15  $S = \{v_4, v_5, v_6\}$  is a neighborhood triple connected dominating set.

12) Any path with a pendant edge attached at each vertex as shown in figure 2.16 is called **Hoffman tree** and is denoted by  $P_n^+$ .

For the **Hoffman tree**,  $\gamma_{ntc}(P_n^+) = n$ .

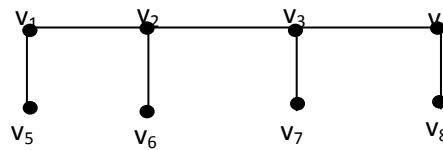


Figure 2.16

In figure 2.16  $S = \{v_5, v_6, v_7, v_8\}$  is a neighborhood triple connected dominating set.

16) The **Moser spindle** (also called the **Mosers' spindle** or **Moser graph**) is an undirected graph, named after mathematicians Leo Moser and his brother William, with seven vertices and eleven edges as shown in figure 2.17.

For the **Moser spindle graph**  $G$ ,  $\gamma_{ntc}(G) = 3$ .

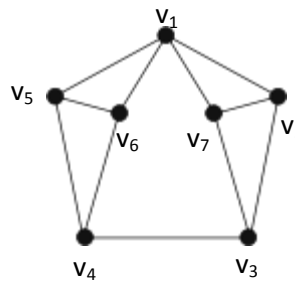


Figure 2.17

Here  $S = \{v_3, v_6\}$  is a neighborhood triple connected dominating set.

**Theorem 2.10** For the path of order  $p \geq 3$ ,  $\gamma_{ntc}(P_p) = \left\lfloor \frac{p}{2} \right\rfloor$ .

**Proof** Let  $P_p = (v_1, v_2, \dots, v_p)$ . If  $p$  is even, then  $S = \{v_i : i = 2k, 2k + 1 \text{ and } k \text{ is odd}\}$  is a ntcd – set of  $P_p$  and if  $p$  is odd, then  $S_1 = S \cup \{v_p\}$  is a ntcd – set of  $P_p$ . Hence  $\gamma_{ntc}(P_p) \leq \left\lfloor \frac{p}{2} \right\rfloor$ . Also if  $S$  is a  $\gamma_{ntc}$  – set of  $P_p$ , then  $N(S)$  contains all the internal vertices of  $P_p$  and hence  $|S| \geq \left\lfloor \frac{p}{2} \right\rfloor$ . Hence  $\gamma_{ntc}(P_p) = \left\lfloor \frac{p}{2} \right\rfloor$ .

**Theorem 2.11** For the cycle of order  $p > 3$ ,  $\gamma_{ntc}(C_p) = \begin{cases} \left\lfloor \frac{p}{2} \right\rfloor & \text{if } p \not\equiv 3 \pmod{4} \\ \left\lfloor \frac{p}{2} \right\rfloor & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

**Proof** Let  $C_p = (v_1, v_2, \dots, v_p, v_1)$ . If  $p = 4k + r$ , where  $0 \leq r \leq 3$ . Let  $S = \{v_i : i = 2j, 2j + 1, j \text{ is odd and } 1 \leq j \leq 2k - 1\}$ . Let  $S_1 = \begin{cases} S & \text{if } p \equiv 0 \pmod{4} \\ S \cup \{v_p\} & \text{if } p \equiv 1 \text{ or } 2 \pmod{4} \\ S \cup \{v_{p-1}\} & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

Clearly  $S_1$  is a ntcd-set of  $C_p$  and hence  $\gamma_{ntc}(C_p) \leq \begin{cases} \left\lfloor \frac{p}{2} \right\rfloor & \text{if } p \not\equiv 3 \pmod{4} \\ \left\lfloor \frac{p}{2} \right\rfloor & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

Now, let  $S$  be any  $\gamma_{ntc}$ -set of  $C_p$ . Then  $\langle S \rangle$  contains at most one isolated vertex and  $\langle N(S) \rangle = \begin{cases} P_{p-1} & \text{if } p \not\equiv 3 \pmod{4} \\ C_p & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

Hence  $\langle S \rangle \geq \begin{cases} \left\lfloor \frac{p}{2} \right\rfloor & \text{if } p \not\equiv 3 \pmod{4} \\ \left\lfloor \frac{p}{2} \right\rfloor & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

The results follows.

**Theorem 2.12** For any connected graph  $G$  with  $p \geq 3$ , we have  $\left\lfloor \frac{p}{\Delta+1} \right\rfloor \leq \gamma_{ntc}(G) \leq p - 1$  and the bounds are sharp.

**Proof** Since any neighborhood triple connected dominating set is a dominating set,  $\left\lfloor \frac{p}{\Delta+1} \right\rfloor \leq \gamma(G) \leq \gamma_{ntc}(G)$ , the lower bound is attained. Also for a connected graph clearly  $\gamma_{ntc}(G) \leq p - 1$ . For  $K_{2,3}$ , the lower bound is attained and for  $P_3$  the upper bound is attained.

**Theorem 2.13** For any connected graph  $G$  with  $p \geq 3$  vertices,  $\gamma_{ntc}(G) = p - 1$  if and only if  $G$  is isomorphic to  $P_3$ ,  $C_3$  and any one of the graphs  $G_1$  and  $G_2$  shown in figure 2.18.

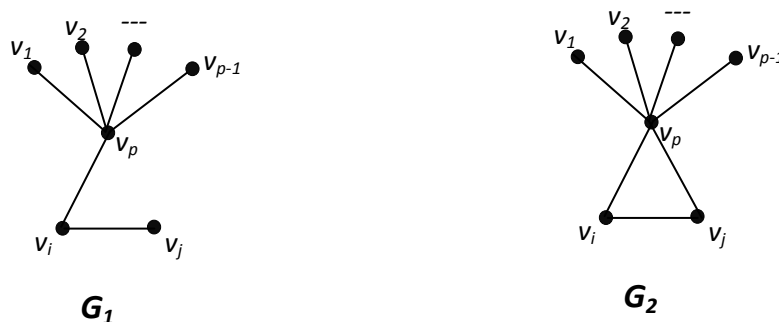


Figure 2.18

**Proof** Suppose  $G$  is isomorphic to any one of the graphs as stated in the theorem, then clearly,  $\gamma_{ntc}(G) = p - 1$ . Conversely, assume that  $G$  is a connected graph with  $p \geq 3$  vertices and  $\gamma_{ntc}(G) = p - 1$ . Let  $S = \{v_1, v_2, \dots, v_{p-1}\}$  be a  $\gamma_{ntc}(G)$ -set. Let  $\langle V - S \rangle = \{v_p\}$ . Since  $S$  is the neighborhood triple connected dominating set, there exists  $v_i$  in  $S$  such that  $v_i$  is adjacent to  $v_p$ . Also  $\langle N(S) \rangle$  is triple connected, we have  $v_i$  is adjacent to  $v_j$  for  $i \neq j$  in  $S$ .

**Case (i)**  $|N(S)| = 3$ .

Then the induced subgraph  $\langle N(S) \rangle$  has the following possibilities.  $\langle N(S) \rangle = P_3$  or  $C_3$ . Hence  $G$  is isomorphic to  $P_3$ ,  $C_3$  or the graphs  $G_1$  and  $G_2$  in figure 2.18.

**Case (ii)**  $|N(S)| > 3$ .

Then there exists atleast one  $v_k$  for  $i \neq j \neq k$  in  $S$  which is adjacent to either  $v_i$  or  $v_j$ .

If  $v_k$  is adjacent to  $v_i$ , then the induced subgraph  $\langle N(S) \rangle$  contains  $K_{1,3}$  and hence it is not triple connected. If we increase the degrees of the vertices in  $S$  we can find a neighborhood triple connected dominating set with fewer elements than  $S$ . Hence no graph exists in this case.

If  $v_k$  is adjacent to  $v_j$ , then we can remove  $v_k$  from  $S$  and find a neighborhood triple connected dominating set of  $G$  with less than  $p - 1$  vertices, which is a contradiction. Hence no graph exists in this case.

**The Nordhaus – Gaddum type result is given below:**

**Theorem 2.14** Let  $G$  be a graph such that  $G$  and  $\bar{G}$  have no isolates of order  $p \geq 3$ . Then

- (i)  $\gamma_{nc}(G) + \gamma_{nc}(\overline{G}) \leq 2p - 2$
- (ii)  $\gamma_{nc}(G) \cdot \gamma_{nc}(\overline{G}) \leq (p - 1)^2$ .

**Proof** The bound directly follows from *theorem 2.12*. For  $K_3$ , both the bounds follows.

### III. RELATION WITH OTHER GRAPH THEORETICAL PARAMETERS

**Theorem 3.1** For any connected graph  $G$  with  $p \geq 3$  vertices,  $\gamma_{nc}(G) + \kappa(G) \leq 2p - 2$  and the bound is sharp if and only if  $G \cong K_3$ .

**Proof** Let  $G$  be a connected graph with  $p \geq 3$  vertices. We know that  $\kappa(G) \leq p - 1$  and by *theorem 2.12*,  $\gamma_{nc}(G) \leq p - 1$ . Hence  $\gamma_{nc}(G) + \kappa(G) \leq 2p - 2$ . If  $G \cong K_3$  then clearly  $\gamma_{nc}(G) + \kappa(G) = 2p - 2$ . Let  $\gamma_{nc}(G) + \kappa(G) = 2p - 2$ . This is possible only if  $\gamma_{nc}(G) = p - 1$  and  $\kappa(G) = p - 1$ . But  $\kappa(G) = p - 1$ , and so  $G \cong K_p$  for which  $\gamma_{nc}(G) = 1$  for  $p > 3$ . Hence  $p = 3$  so that  $G \cong K_3$ .

**Theorem 3.2** For any connected graph  $G$  with  $p \geq 3$  vertices,  $\gamma_{nc}(G) + \Delta(G) \leq 2p - 2$  and the bound is sharp.

**Proof** Let  $G$  be a connected graph with  $p \geq 3$  vertices. We know that  $\Delta(G) \leq p - 1$  and by *theorem 2.12*,  $\gamma_{nc}(G) \leq p - 1$ . Hence  $\gamma_{nc}(G) + \Delta(G) \leq 2p - 2$ . For  $K_3$ , the bound is sharp.

**Theorem 3.3** For any connected graph  $G$  with  $p \geq 3$  vertices,  $\gamma_{nc}(G) + \chi(G) \leq 2p - 1$  and the bound is sharp if and only of  $G \cong K_3$ .

**Proof** Let  $G$  be a connected graph with  $p \geq 3$  vertices. We know that  $\chi(G) \leq p$  and by *theorem 2.12*,  $\gamma_{nc}(G) \leq p - 1$ . Hence  $\gamma_{nc}(G) + \chi(G) \leq 2p - 1$ . Suppose  $G \cong K_3$ , then clearly  $\gamma_{nc}(G) + \chi(G) = 2p - 1$ . Let  $\gamma_{nc}(G) + \chi(G) = 2p - 1$ . This is possible only if  $\gamma_{nc}(G) = p - 1$  and  $\chi(G) = p$ . Since  $\chi(G) = p$ ,  $G$  is isomorphic to  $K_p$  for which  $\gamma_{nc}(G) = 1$ , for  $p > 3$ . Hence  $p = 3$ , so that  $G \cong K_3$ .

### IV. RELATION WITH OTHER DOMINATION PARAMETERS

**Theorem 4.1** For any connected graph  $G$  with  $p \geq 5$  vertices,  $\gamma_{tc}(G) + \gamma_{nc}(G) < 2p - 3$ .

**Proof** Let  $G$  be a connected graph with  $p \geq 5$  vertices. We know that  $\gamma_{tc}(G) \leq p - 2$  for  $p \geq 5$  vertices and by *theorem 2.12*,  $\gamma_{nc}(G) \leq p - 1$  for  $p \geq 3$  vertices. Hence  $\gamma_{tc}(G) + \gamma_{nc}(G) \leq 2p - 3$  for  $p \geq 5$  vertices. Also by *theorem 2.13*, the bound is not sharp.

**Theorem 4.2** For any connected graph  $G$  with  $p \geq 4$  vertices,  $\gamma_{ctc}(G) + \gamma_{nc}(G) < 2p - 4$ .

**Proof** Let  $G$  be a connected graph with  $p \geq 4$  vertices. We know that  $\gamma_{ctc}(G) \leq p - 3$  for  $p \geq 4$  vertices and by *theorem 2.12*,  $\gamma_{nc}(G) \leq p - 1$  for  $p \geq 3$  vertices. Hence  $\gamma_{ctc}(G) + \gamma_{nc}(G) \leq 2p - 4$  for  $p \geq 4$  vertices. Also by *theorem 2.13*, the bound is not sharp.

**Theorem 4.3** For any connected graph  $G$  with  $p \geq 5$  vertices,  $\gamma_{cptc}(G) + \gamma_{nc}(G) < 2p - 3$ .

**Proof** Let  $G$  be a connected graph with  $p \geq 5$  vertices. We know that  $\gamma_{cptc}(G) \leq p - 2$  for  $p \geq 5$  vertices and by *theorem 2.12*,  $\gamma_{nc}(G) \leq p - 1$  for  $p \geq 3$  vertices. Hence  $\gamma_{cptc}(G) + \gamma_{nc}(G) \leq 2p - 3$  for  $p \geq 5$  vertices. Also by *theorem 2.13*, the bound is not sharp.

### REFERENCES

- [1] G. Mahadevan, A. Selvam, J. Paulraj Joseph and, T. Subramanian, *Triple connected domination number of a graph*, International Journal of Mathematical Combinatorics, Vol.3 (2012), 93-104.
- [2] G. Mahadevan, A. Selvam, J. Paulraj Joseph, B. Ayisha and, T. Subramanian, *Complementary triple connected domination number of a graph*, Advances and Applications in Discrete Mathematics, Vol. 12(1) (2013), 39-54.
- [3] G. Mahadevan, A. Selvam, A. Mydeen bibi and, T. Subramanian, *Complementary perfect triple connected domination number of a graph*, International Journal of Engineering Research and Application, Vol.2, Issue 5 (2012) , 260-265.
- [4] G. Mahadevan, A. Selvam, A. Nagarajan, A. Rajeswari and, T. Subramanian, *Paired Triple connected domination number of a graph*, International Journal of Computational Engineering Research, Vol. 2, Issue 5 (2012), 1333-1338.

- [5] G. Mahadevan, A. Selvam, B. Ayisha, and, T. Subramanian, *Triple connected two domination number of a graph*, International Journal of Computational Engineering Research Vol. 2, Issue 6 (2012),101-104.
- [6] G. Mahadevan, A. Selvam, V. G. Bhagavathi Ammal and, T. Subramanian, *Restrained triple connected domination number of a graph*, International Journal of Engineering Research and Application, Vol. 2, Issue 6 (2012), 225-229.
- [7] G. Mahadevan, A. Selvam, M. Hajmeeral and, T. Subramanian, *Dom strong triple connected domination number of a graph*, American Journal of Mathematics and Mathematical Sciences, Vol. 1, Issue 2 (2012), 29-37.
- [8] G. Mahadevan, A. Selvam, V. G. Bhagavathi Ammal and, T. Subramanian, *Strong triple connected domination number of a graph*, International Journal of Computational Engineering Research, Vol. 3, Issue 1 (2013), 242-247.
- [9] G. Mahadevan, A. Selvam, V. G. Bhagavathi Ammal and, T. Subramanian, *Weak triple connected domination number of a graph*, International Journal of Modern Engineering Research, Vol. 3, Issue 1 (2013), 342-345.
- [10] G. Mahadevan, A. Selvam, N. Ramesh and, T. Subramanian, *Triple connected complementary tree domination number of a graph*, International Mathematical Forum, Vol. 8, No. 14 (2013), 659-670.
- [11] G. Mahadevan, B. Anitha, A. Selvam, and, T. Subramanian, *Efficient complementary perfect triple connected domination number of a graph*, Journal of Ultra Scientist of Physical Sciences, Vol. 25(2)A (2013), 257-268.
- [12] G. Mahadevan, N. Ramesh, A. Selvam, and, T. Subramanian, *Efficient triple connected domination number of a graph*, International Journal of Computational Engineering Research, Vol. 3, Issue 6 (2013), 1-6.
- [13] G. Mahadevan, N. Ramesh, A. Selvam, S. Rajeswari and, T. Subramanian, *Complementary triple connected clique domination number of a graph*, Accepted for publication in Journal of Ultra Scientist of Physical Sciences, Vol. 25. No.(3) (2013) - Preprint.
- [14] G. Mahadevan, M. Lavanya, A. Selvam, T. Subramanian, *Triple connected.com domination number of a graph*, International Journal of Mathematical Archive, Vol. 4(10) (2013), 178-189.
- [15] J. A. Bondy and U. S. R. Murty, *Graph Theory*, Springer, 2008.
- [16] J. Paulraj Joseph and, S. Arumugam, *Domination and connectivity in graphs*, International Journal of Management Systems, 8 (3) (1992), 233–236.
- [17] J. Paulraj Joseph and, S. Arumugam, *Domination and coloring in graphs*, International Journal of Management Systems, 8 (1) (1997), 37–44.
- [18] J. Paulraj Joseph, M. K. Angel Jebitha, P. Chithra Devi and, G. Sudhana, *Triple connected graphs*, Indian Journal of Mathematics and Mathematical Sciences, Vol. 8, No.1 (2012), 61-75.
- [19] J. Paulraj Joseph and, G. Mahadevan, *On complementary perfect domination number of a graph*, Acta Ciencia Indica, Vol. XXXI M, No. 2. (2006), 847–853.
- [20] J. Paulraj Joseph and G. Mahadevan, *Cubic graphs with equal domination number and chromatic number*. Proceedings of the National L conference on graph theory and its applications, Narosa publications, India, (2001), 129-139.
- [21] S. Arumugam and C. Sivagnanam, *Neighborhood Connected Domination in Graphs*, J. Combin. Math. Combin. Comp., 73 (2010), 55-64.
- [22] S. Arumugam and C. Sivagnanam, *Neighborhood Connected Domatic number of a Graph*, J. Combin. Math. Combin. Comp., 75 (2010), 239-249.
- [23] S. Arumugam and C. Sivagnanam, *Neighborhood Total Domination in Graphs*, Opuscula Mathematica, Vol. 31, No.4 (2011), 519-531.
- [24] T. W. Haynes, S. T. Hedetniemi and, P. J. Slater, *Domination in graphs*, Advanced Topics, Marcel Dekker, New York (1998).
- [25] T. W. Haynes, S. T. Hedetniemi and, P. J. Slater, *Fundamentals of domination in graphs*, Marcel Dekker, New York (1998).

## Bioelectrical Impedance Analysis (BIA) For Assessing Tbw and Ffm of Indian Females

<sup>1</sup>. Munna Khan, <sup>2</sup>Shabana Mehfuz, <sup>3</sup>. Ghazala PerveenKhan

<sup>1</sup>Electrical Engineering Department, Jamia Millia Islamia, New Delhi-110025

<sup>2</sup>Professor, Electrical Engineering Department, Jamia Millia Islamia, New Delhi-110025

<sup>3</sup>Assistant Professor, Electrical Engineering Department, Jamia Millia Islamia, New Delhi-110025

### ABSTRACT

**Background:** The Bioelectrical Impedance analysis is an easy, applicable method for assessing Total Body water and Fat Free Mass of various groups of people. It has many advantages over other methods and is safe, rapid, portable, easy to perform and require minimum operator training. It has been used extensively for developing specific prediction equation for different ethnicity, age, gender, level of body fatness and physical activity. Regression equations play great role to estimate the body density and fatness specific to the owing to methodological and biological factors.

**Purpose:** The purpose of the study was to investigate the utility of multi-frequency BIA for the estimation of TBW and FFM of Indian females at different frequencies. Earlier scientists have measured various parameters of body composition of Indian population using MALTRON-II. However, literature shows that prediction equations have not been developed for Indian females which could develop the healthy prediction equation for Indian females (unhealthy eating habits). It has been found in the study that women particularly Indian womens have unhealthy eating habits owing to the fact that they concentrate only on 1 aspect was their food cooked good enough to eat, but they don't understand the meaning of healthy food. In the present research paper, an attempt has been made to develop BIA equations using data taken from PhD thesis (1). This data was taken from my senior Dr. Goswami, who collected the data in the college in which he was teaching. Some of the data that my supervisor, co-supervisor and myself has taken in DRDO have already been utilized in building multi-compartmental model, developing generalized age and sex specific prediction equation and developing REE of Indian subjects.

**Keywords:** R(2.9.2) software, Bio Electrical Impedance Analysis, Prediction Equation, MALTRON-II, Multiple Regression Analysis, Total Body Water, Fat Free Mass, Impedance index.

**Methods:** Data of reference (1) included vital information about body composition of 60 Indian females such as age, TBW, FFM, height, weight, impedance and phase at 5KHz., 50KHz., 100KHz. and 200 KHz.. This data was then used to calculate the impedance index i.e. ( $\text{height}^2/\text{impedance}$ ) at the respective frequencies. This method uses multiple regression analysis to drive prediction equation of Indian females at frequencies of 5KHz.,50KHz.,100KHz.,200KHz., As we Know that in multiple regression analysis wherein the main aim is to predict dependent variable from the no. of independent variables that are known which is called dependent variable from 1 or more independent variable also called as predictor variable. One more theme in developing prediction equation of Here in this study we used TBW and FFM as dependent variable and Weight of females and Impedance index i.e.( $\text{height}^2/\text{impedance}$ ) as independent variable at different frequencies of 5 KHz.,50KHz.,100KHz.,200KHz respectively. Data was fed then into R software, an integrated suite of software facilities used for data manipulation, calculation and graphical display. It is used for effective data handling and storage facility. The main use of R software is for regression analysis, which was then used to generate the prediction equation. The prediction developed here is through multiple regression analysis. In multiple regression analysis the equation is of the form given below;

$Y = m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_nx_n + c$ ; where c is intercept and  $m_1, m_2, m_3, \dots, m_n$  are the weight's assigned to each of the predictor variables by the regression solution.

**Result:** Data used in commercial software provided 8 BIA equations; 4 for TBW and 4 for FFM at frequencies of 5 KHz, 50KHz, 100 KHz and 200 KHz. And so 8 sets of dependent variable were there, This data included other statistical data such as mean, standard deviation and correlation of Total Body Water (TBW) and Fat Free Mass (FFM) with Impedance index and Weight. Besides this scatter matrix plot for Total Body Water (TBW) and Fat free Mass (FFM), Normal distribution of standardised residuals showing the relationship between TBW

and FFM ,Scale location plot, Residual verses leverage plot, standardised residual verses cook’s distance plot at frequencies of 5 KHz,50KHz,100 KHz and 200 KHz are plotted. These plots for linear model objects give the diagnostic information about the linear model.

**Conclusions:** The final race-combined TBW prediction equations included stature<sup>2</sup>/resistance and body weight. Multiple regression analysis was carried out on clinical data through R 2.9 software. The BIA prediction equation for Total Body water and Fat Free mass was developed at different frequencies of 5 KHz. and 50 KHz. respectively. The data was taken for Indian females lying in a limited age span of 17-22 years.

**I. INTRODUCTION:**

Developing prediction equation for Indian females for evaluating and monitoring growth and nutritional status is an important area of research. Assessing the body composition through isotope dilution, hydro densitometry, dual X-ray absorptiometry (DXA),air displacement plethysmography, magnetic resonance imaging, are sometimes used for body composition analysis, but these equipments are not easily available and is expensive to maintain, so their use in clinical and field studies is limited. One of the most popular methods for body composition analysis is through Bioelectrical Impedance Analysis. One of the main advantages of using BIA method is that it does not require high technician skill, it is generally more comfortable and does not intrude much on client’s privacy. Traditional BIA method which is still most frequently used involves the measurement of impedance at single frequency generally at 50KHz. Although single frequency BIA is most used in clinical practices, this device could not predict Total Body Water accurately. MF-BIA seems to give a better estimation of hydration than SF-BIA because the principle of measuring the flow of current through the body (impedance) is dependent on the frequency applied. At low frequencies, the current cannot bridge the cellular membrane and will pass predominantly through the extracellular space. At higher frequencies penetration of the cell membrane occurs and the current is conducted by both the extra-cellular water (ECW) and intra-cellular water (ICW).

In India when it comes to general health of people there is a large disparity between urban elite class and rural class. Obesity is becoming a factor in many nations around the world. According to latest obesity statistics, sponsored by International Day of Evaluation of Abdominal Obesity;75 percent of Indian women and 58 percent of Indian men are obese. Besides this, numerous studies indicate that malnutrition is another serious health concern that Indian women face (Chatterjee, 1990; Desai, 1994; The World Bank, 1996). It threatens their survival as well as that of their children. The negative effects of malnutrition among women are compounded by heavy work demands, by poverty, by childbearing and rearing, and by special nutritional needs of women, resulting in increased susceptibility to illness and consequent higher mortality. Attention, must therefore be paid to determine the body composition of females so that appropriate measures can be taken if women in India are facing abnormality in their health due to their abnormal nutritional status.

The purpose of the current study was to use the female body composition database to develop and predict TBW and FFM at the frequencies of 5KHz, 50KHz, 100KHz, 200KHz so that, a general idea about the health status of Indian female can be observed and for that purpose multiple regression analysis is done.

**Subjects and Methods:** Data of 60 Indian females from reference(1) was used to develop and predict TBW and FFM at 5KHz,50KHz,100KHz,200KHz which was then used for the descriptive analysis of Indian females which included their age, weight, height, Total Body Water(TBW), Fat Free Mass (FFM), and Impedance at these frequencies. These data were then used to calculate the body stature. i.e. (height<sup>2</sup>/impedance) of females. In this paper, a powerful statistical program R (version 2.9.2); basically used for statistical analysis is used. Here we have used this software to develop linear model for Total Body Water(TBW) and Fat Free Mass(FFM) at different frequency and form a prediction equation using weight and Impedance index i.e. (height<sup>2</sup>/impedance) of females at these frequencies as independent variables and TBW and FFM as dependent variables. The obtained equations are of the form:

$TBW_{(f1,f2,f3,f4)}=a_0Wt.+a_1Z_{I(f1,f2,f3,f4)} +C_{1(f1,f2,f3,f4)}.....(1)$

$FFM_{(f1,f2,f3,f4)}=b_0Wt.+b_1Z_{I(f1,f2,f3,f4)} +C_{2(f1,f2,f3,f4)}.....(2)$

Where  $TBW_{(f1,f2,f3,f4)}$  and  $FFM_{(f1,f2,f3,f4)}$  is Total Body Water and Fat Free Mass at frequencies  $f1=5KHz.,f2=50KHz.,f3=100KHz.,f4=200KHz.$  and  $Z_{I(f1,f2,f3,f4)}$  is the calculated Impedance index i.e. (height<sup>2</sup>/impedance) of females at these frequencies.  $C_{1(f1,f2,f3,f4)}$  and  $C_{2(f1,f2,f3,f4)}$  are intercepts of equation (1) and (2) respectively and  $a_0, b_0$  are coefficients multiplied by weight variable of equation(1) and(2) respectively and  $a_1, b_1$  are coefficients multiplied by Impedance index variable of equation(1) and(2) respectively. The flowchart showing the sequence of operations of algorithm used in R 2.9 software to develop and predict linear model of TBW, FFM at these multiple frequencies is also shown in Figure(1)



**TABLE 1.**

Clinical data of 60 Indian females from reference (1), showing age, weight, sex and customer ID of subjects

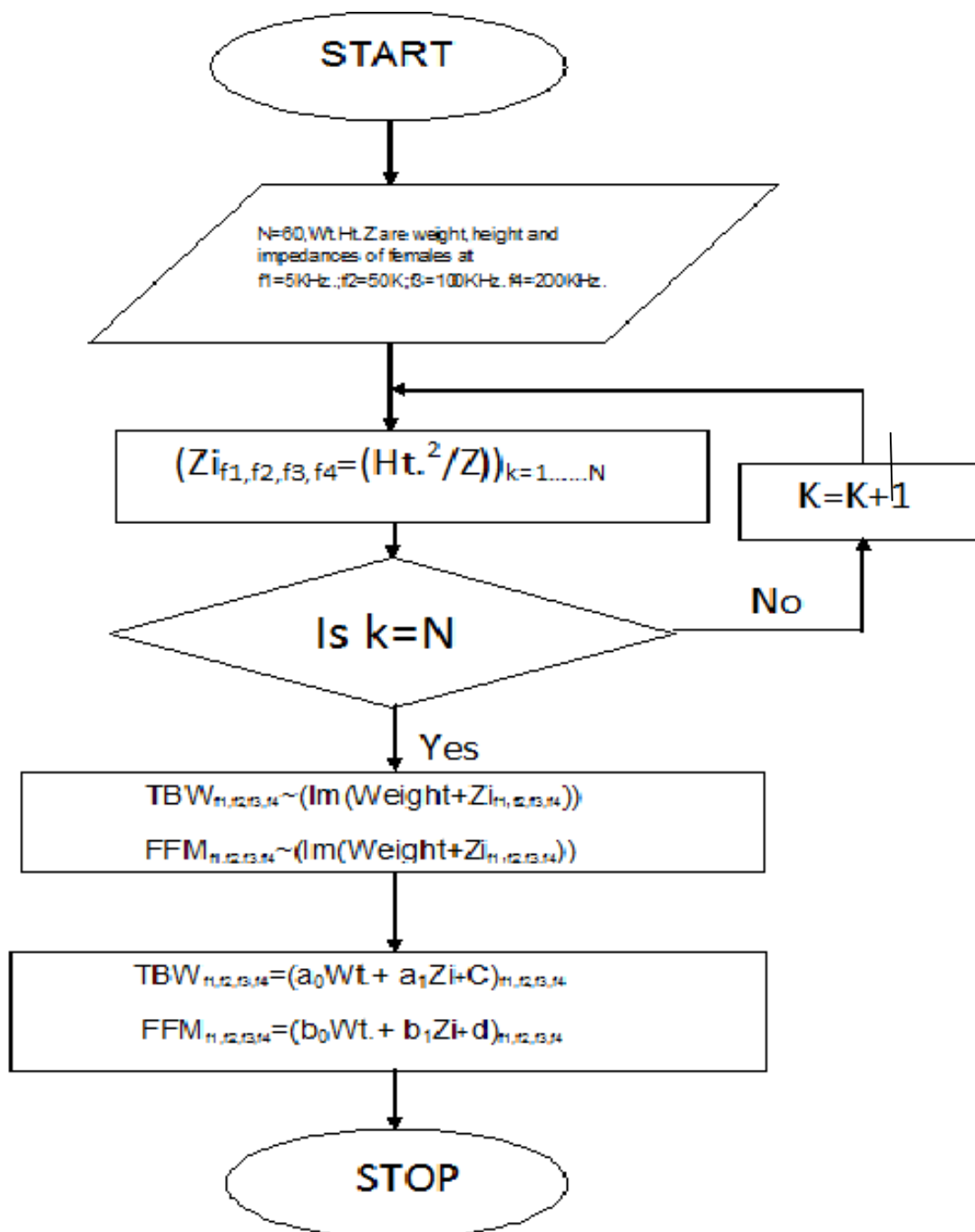
Serial no.	Customer Id	Weight(Kg)	Sex(Male)	Age(Years)
1.	187	44	1	17
2.	127	50	1	17
3.	173	45	1	17
4.	108	53	1	18
5.	140	46	1	18
6.	167	62	1	18
7.	169	52	1	18
8.	171	39	1	18
9.	176	52	1	18
10.	183	57	1	18
11.	186	44	1	18
12.	188	51	1	18
13.	189	39	1	18
14.	190	43	1	18
15.	191	54	1	18
16.	192	50	1	18
17.	193	58	1	18
18.	203	40	1	18
19.	206	54	1	18
20.	195	57	1	18
21.	208	55	1	18
22.	207	51	1	19
23.	115	50	1	19
24.	163	56	1	19
25.	164	63	1	19
26.	165	48	1	19
27.	168	57	1	19
28.	170	45	1	19
29.	172	50	1	19
30.	177	56	1	19
31.	178	60	1	19
32.	181	51	1	19
33.	182	54	1	19
34.	184	60	1	19
35.	211	60	1	19
36.	212	61	1	19
37.	214	52	1	19
38.	201	52	1	19
39.	166	56	1	20
40.	194	47	1	21
41.	175	37	1	21
42.	180	52	1	22
43.	185	47	1	22
44.	209	53	1	22
45.	71	39	1	23
46.	179	50	1	23
47.	210	58	1	23
48.	54	41	1	24
49.	119	42	1	24
50.	220	45	1	25
51.	53	47	1	25
52.	51	54	1	26
53.	118	62	1	30
54.	124	48	1	30
55.	144	72	1	31
56.	162	55	1	32
57.	125	57	1	33
58.	174	56	1	37
59.	92	50	1	41
60.	197	77	1	42

**TABLE 2.**

Clinical data of 60 Indian females from reference(1) showing impedance and impedance index at frequencies of 5KHz.,50KHz.,100KHz.,200KHz.

S.No.	FFM(Kg)	TBW(Lt.)	Z at 5KHz.	Z at 50 KHz.	Z at 100 KHz.	Z at 200 KHz.	Z <sub>i</sub> at 5KHz.	Z <sub>i</sub> at 50 KHz.	Z <sub>i</sub> at 100KHz.	Z <sub>i</sub> at 200KHz.
1.	37.61	24.35	845	769	718	718	29.17	32.05	34.33	34.33
2.	38.38	26.22	793	692	654	619	29.135	33.387	35.327	37.32
3.	24.02	11.82	871	827	762	762	28.66	30.186	32.76	32.76
4.	43.27	27.79	861	766	725	692	31.33	36.41	38.47	40.31
5.	39.23	25.28	890	792	755	723	32.64	33.136	34.76	36.29
6.	45.91	30.19	804	731	682	682	35.104	38.61	41.384	41.384
7.	39.9	25.81	869	794	731	731	29.09	31.84	34.58	34.58
8.	34.87	23.27	843	761	723	723	28.499	31.57	34.518	33.229
9	38.31	25.39	890	816	757	757	26.99	29.44	31.737	31.737
10.	49.24	27.38	809	739	696	696	30.857	33.781	35.868	35.868
11.	36.7	23.06	954	873	846	846	26.167	28.59	29.508	29.508
12.	39.2	26.04	765	720	674	674	31.405	33.37	35.645	35.645
13.	33.54	21.64	936	872	819	819	24.634	26.495	28.21	28.21
14.	36.59	24.03	869	775	728	728	27.646	31	33.001	33.001
15.	39.86	25.73	853	804	760	760	29.266	31.05	32.8473	32.84736.8
16.	41.94	26.24	887	785	748	748	39.066	35.103	6.84	4
17.	40.56	27.16	881	793	723	723	28.335	31.48	34.53	34.53
18.	32.93	21.6	942	835	780	780	22.939	25.88	27.71	27.71
19.	38.4	25.22	884	838	786	786	27.177	28.669	30.566	30.566
20.	44.04	28.66	827	766	722	722	33.72	36.41	38.63	38.63
21.	39.88	25.85	891	821	774	774	28.02	30.41	32.25	32.25
22.	36.95	24.21	943	860	809	809	24.83	27.22	28.94	28.94
23.	40.89	25.93	953	839	798	763	28.56	32.45	34.117	35.68
24.	43.19	27.51	842	783	737	737	32.73	35.19	37.389	37.389
25.	43.77	29.13	828	743	707	707	32.088	35.76	37.58	37.58
26.	38.47	24.35	878	814	759	759	28.43	30.668	32.89	32.89
27.	39.25	27.25	792	723	681	681	29.56	32.37	34.37	34.37
28.	34.56	23.13	879	805	763	763	24.919	27.21	28.71	28.71
29.	39.88	26.08	822	764	725	725	30.755	33.09	34.87	34.87
30.	41.2	28.81	713	624	581	581	33.26228.	38.006	40.82	40.82
31.	40.13	27.62	865	766	724	724	134	31.77	33.613	33.613
32.	41.18	26.24	893	809	771	771	30.118	33.246	34.884	34.884
33.	40.16	26.62	829	770	717	717	30.113	32.42	34.817	34.817
34.	45.89	28.9	875	801	755	755	33.81	36.93	39.184	39.184
35.	44	28.3	868	783	740	740	31.75	35.193	37.24	37.24
36.	45.1	29.01	833	769	740	740	33.88	36.7	38.14	38.14
37.	41.25	25.35	1004	910	843	843	27.78	30.65	33.08	33.08
38.	39.17	25	850	857	810	810	29.74	29.49	31.21	31.21
39.	39.08	26.32	900	815	769	769	27.04	29.86	31.646	31.646
40.	37.06	23.51	983	911	851	851	25.39	27.402	29.335	29.335
41.	31.91	20.75	1100	1006	938	938	21.56	23.57	25.28	25.28
42.	37.91	25.85	833	761	717	717	28.102	30.76	32.648	32.648
43.	6.3	25.49	870	825	778	778	32.83	34.62	36.72	36.72
44.	12.33	25.99	873	808	765	765	30.06	32.48	34.305	34.305
45.	7.19	21.06	991	903	857	842	22.1	24.256	25.56	26.014
46.	12.34	25.74	803	712	668	668	28.772	32.45	34.586	34.586
47.	12.73	25.7	892	827	768	768	25.224	27.21	29.29	29.29
48.	5.29	23.07	932	842	803	804	27.812	30.785	32.28	32.24
49.	5.6	23.88	967	855	813	781	27.476	31.07	32.68	34.019
50.	16.6	25.96	966	876	819	819	22.98	25.34	27.104	27.104
51.	12.49	23.5	884	805	760	749	24.77	27.21	28.82	29.24
52.	13.66	26.21	920	829	785	775	28.526	31.65	33.43	33.43
53.	21.16	29.6	730	645	613	584	32.911	37.248	39.19	41.138
54.	10.66	26.25	800	666	626	592	28.88	34.69	36.91	39.03
55.	30.27	31.27	819	705	665	635	30.48	35.41	37.54	39.313
56.	16.6	25.96	816	745	702	702	30.59	33.51	35.56	35.56
57.	16.53	29.11	752	631	591	558	31.948	38.074	40.65	43.05
58.	15.82	27.61	803	743	691	691	31.88	34.45	30.18	37.05
59.	11.07	26.02	871	780	739	722	30.13	33.65	35.512	35.512
60.	36.2	32.48	708	639	605	605	34.815	38.57	40.74	40.74

Flowchart showing the general process carried to develop and predict TBW and FFM at different frequencies:



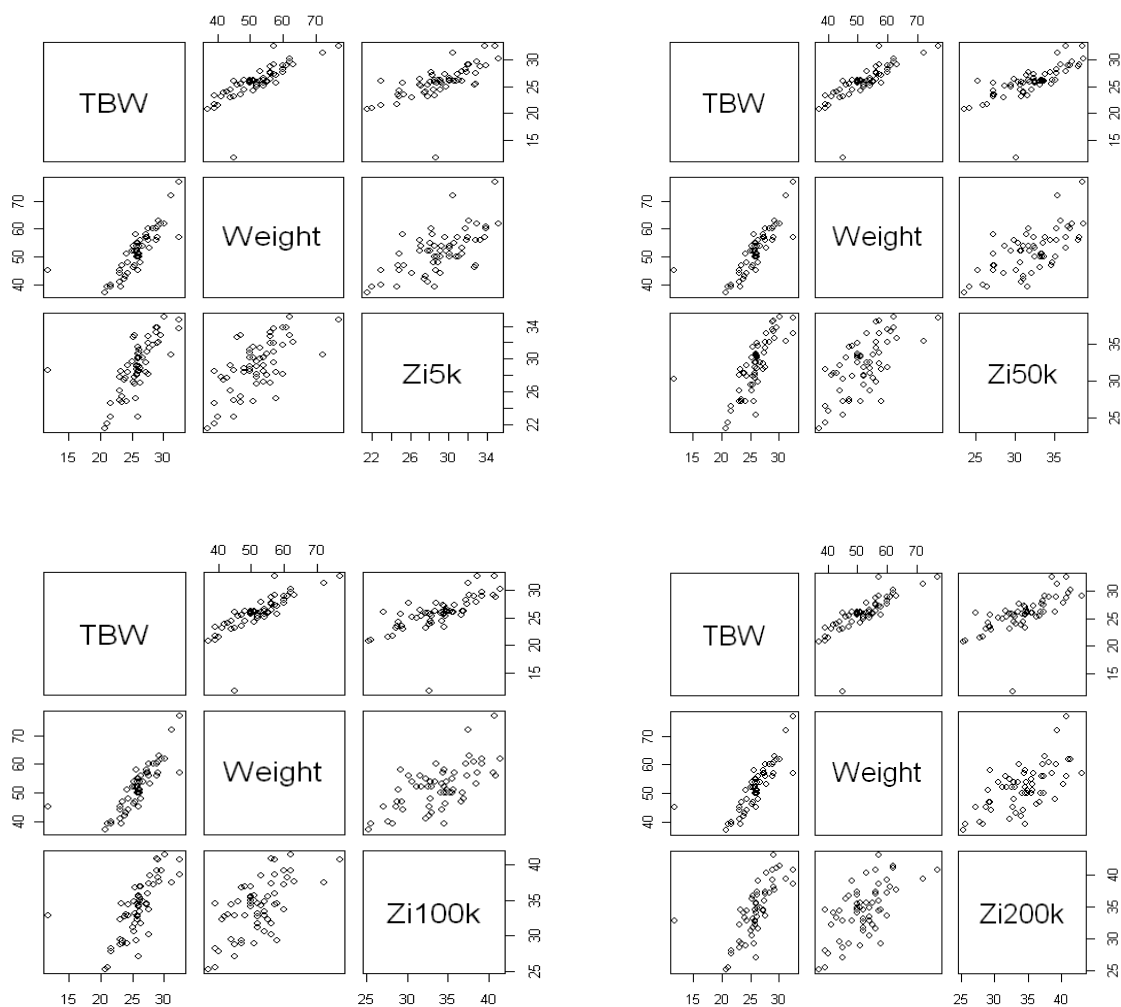
**Fig1:** Flowchart showing the general process to develop linear model of TBW and FFM at frequencies of 5KHz, 50KHz, 100KHz and 200KHz

**Results and Statistical Analysis:** The equation used to predict TBW and FFM were developed at the frequencies of 5KHz, 50KHz, 100KHz, 200KHz. by using the data from table 1 and 2. The table below show the predicted TBW and FFM equation at different frequencies and their statistical analysis:

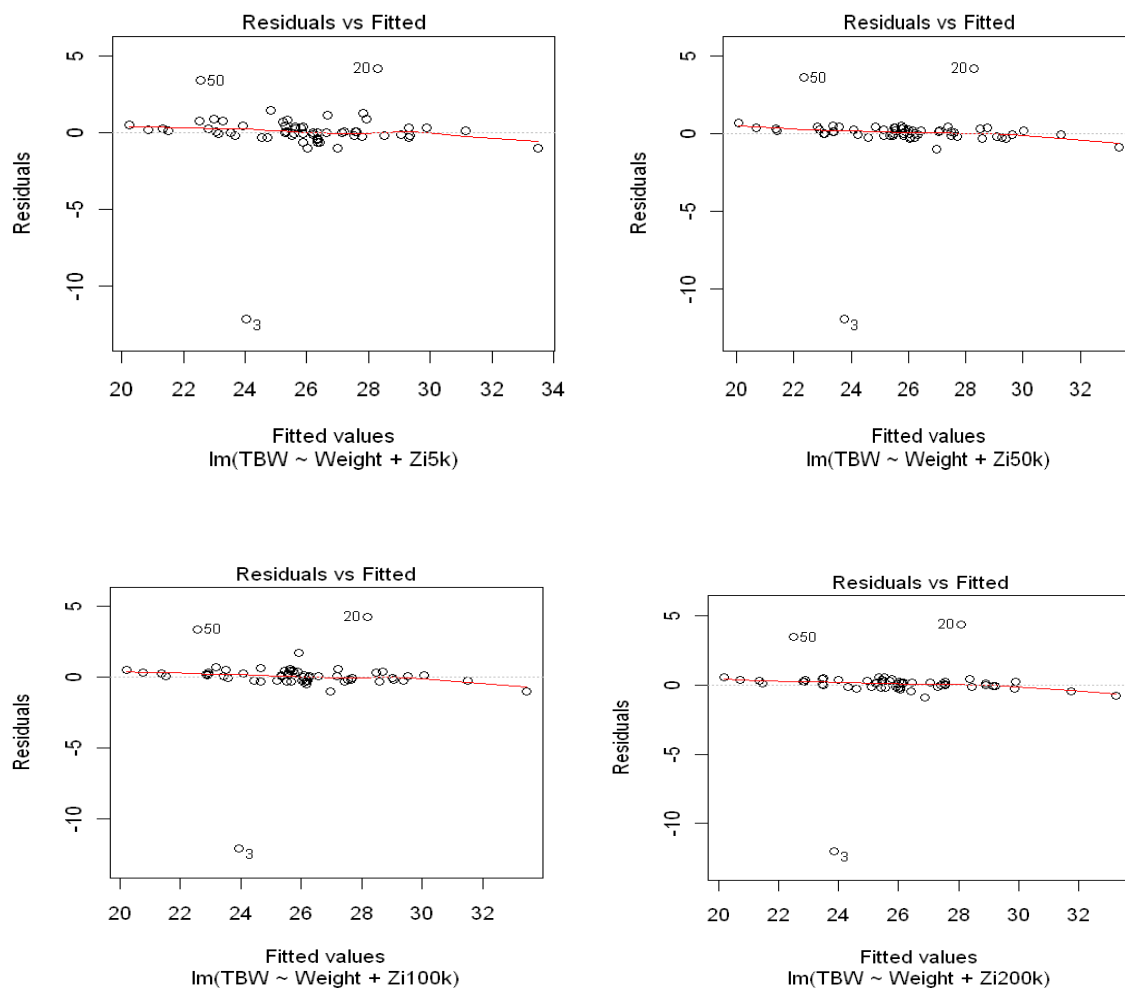
**TABLE 3.**

S.No.	Prediction Equations developed	Frequency	Mean	Standard deviation	Standard error	Residual error	Multiple R <sup>2</sup>	Adjusted R <sup>2</sup>
1.	TBW=0.24689Wt.+0.25593Zi+5.58336	5KHz	TBW=25.85 53 Wt.=51.9333 Zi=29.10972	TBW=3.093 788 Wt.=7.71717 6 Zi=3.152058	Intercept = 2.24234 Wt.=0.0410 2 Zi=0.10042	1.845 on 57 df	0.656 4	0.64 44
2.	TBW=0.22363Wt.+0.9178Zi+4.89741	50K Hz	TBW=25.85 53 Wt.=51.9333 Zi=32.02492	TBW=3.093 788 Wt.=7.71717 6 Zi=3.629973	Intercept=2. 06674 Wt.=0.2236 3 Zi=0.08522	1.773 on 57 df	0.682 6	0.67 14
3.	TBW=0.23751Wt.+0.24204Zi+5.31032	100K Hz	TBW=25.85 53 Wt.=51.9333 Zi=33.921	TBW=3.093 788 Wt.=7.71717 6 Zi=3.837306	Intercept=2. 12356 Wt.=0.0399 6 Zi=0.8057	1.809 on 57 df	0.669 6	0.65 8
4.	TBW=0.24689Wt.+0.25593Zi+5.58336	200K Hz	TBW=25.85 53 Wt.=51.9333 Zi=34.31884	TBW=3.093 788 Wt.=7.71717 6 Zi=4.070613	Intercept=2. 01648 Wt.=0.0392 1 Zi=0.07433	1.787 on 57 df	0.677 6	0.66 63
5.	FFM=0.12496Wt.+0.56707Zi+15.91914	5KHz	FFM=38.916 33 Wt.=51.9333 Zi=29.10988	FFM=3.9643 2 Wt.=7.71717 6 Zi=3.152027	Intercept=3. 7801 Wt.=0.0691 5 Zi=0.16929	3.11 on 57 df	0.405 4	0.38 45
6.	FFM=0.11924Wt.+0.50014Zi+16.70689	50K Hz	FFM=38.916 33 Wt.=51.9333 Zi=32.02492	FFM=3.9643 2 Wt.=7.71717 6 Zi=3.629973	Intercept=3. 62533 Wt.=0.0703 2 Zi=0.14949	3.111 on 57 df	0.405 1	0.38 43
7.	FFM=0.1392Wt.+0.42686Zi+17.20422	100K Hz	FFM=38.916 33 Wt.=51.9333 Zi=33.92158	FFM=3.9643 2 Wt.=7.71717 6 Zi=3.827306	Intercept=3. 70514 Wt.=0.0697 1 Zi=0.14057	3.157 on 57 df	0.387 4	0.36 59
8.	FFM=0.21844Wt.+0.34863Zi+15.78796	200K Hz	FFM=39.109 Wt.=51.9333 Zi=34.35368	FFM=3.9268 Wt.=7.71717 6 Zi=4.160915	Intercept=3. 05025 Wt.=0.0608 1 Zi=0.11278	2.752 on 57 df	0.525 6	0.50 9

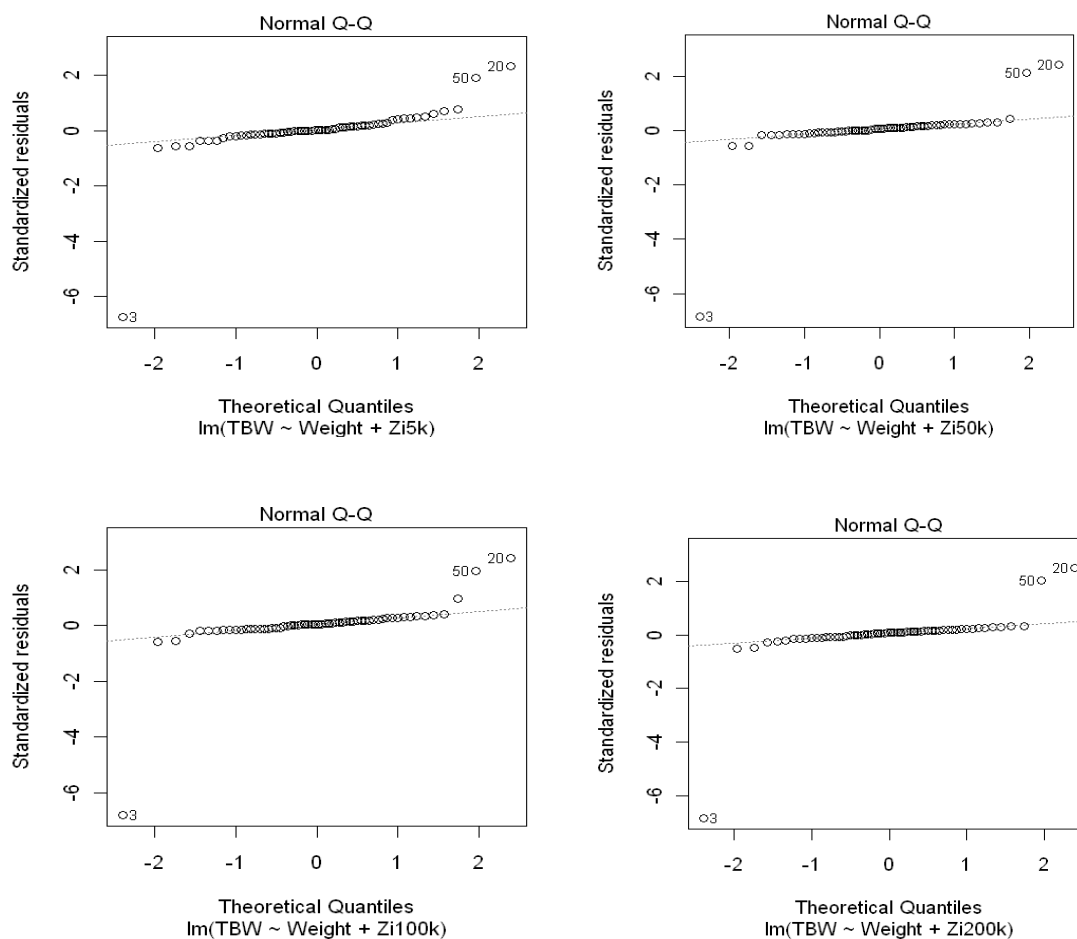
**Graphical Analysis:** The graphical interpretation of data include scatter matrix plot, Random scatter distribution, normal distribution, scale location plot, residual verses leverage plot and standardized residual verses cook's distance plot at different frequencies



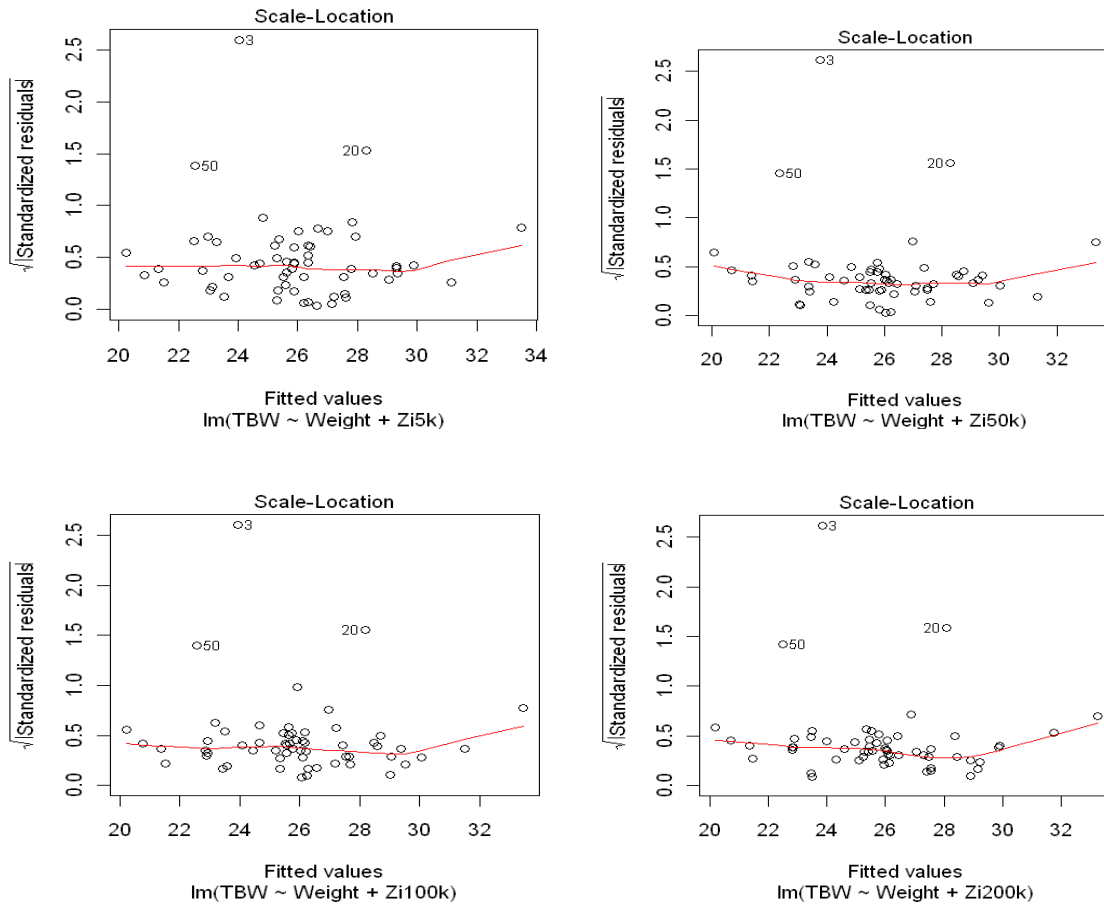
**Fig.2:** Scatter Plot Matrix distribution of body composition of Indian females showing the relationship between Total Body Water(TBW)in (litre) , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at frequencies of 5KHz,50KHz,100KHz,200KHz)in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg.



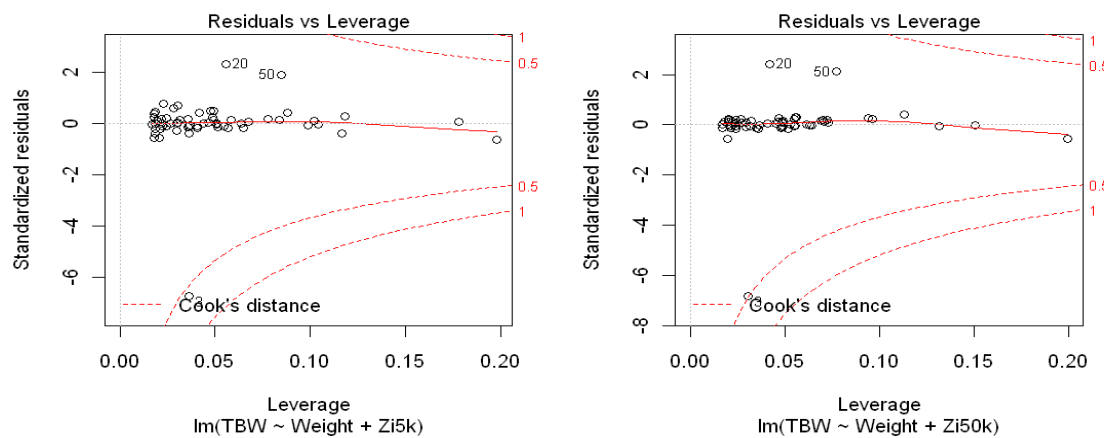
**Fig3** :Random scatter distribution of residual versus fitted values of Indian males showing the relationship between Total Body Water(TBW)in (litre) , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at 5KHz.frequency)in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg. at 5KHz.



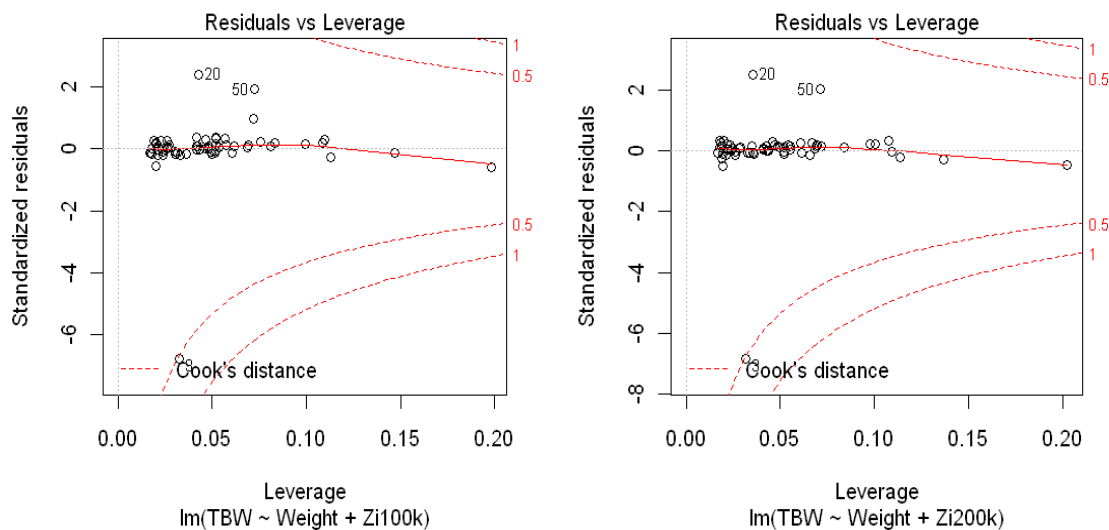
**Fig.4:** Normal distribution versus Standardized residuals of Indian females showing the relationship between Total Body Water(TBW)in (litre) , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at frequencies of 5KHz.,50KHz.,100KHz.,200KHz.)in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg. at 5 KHz.



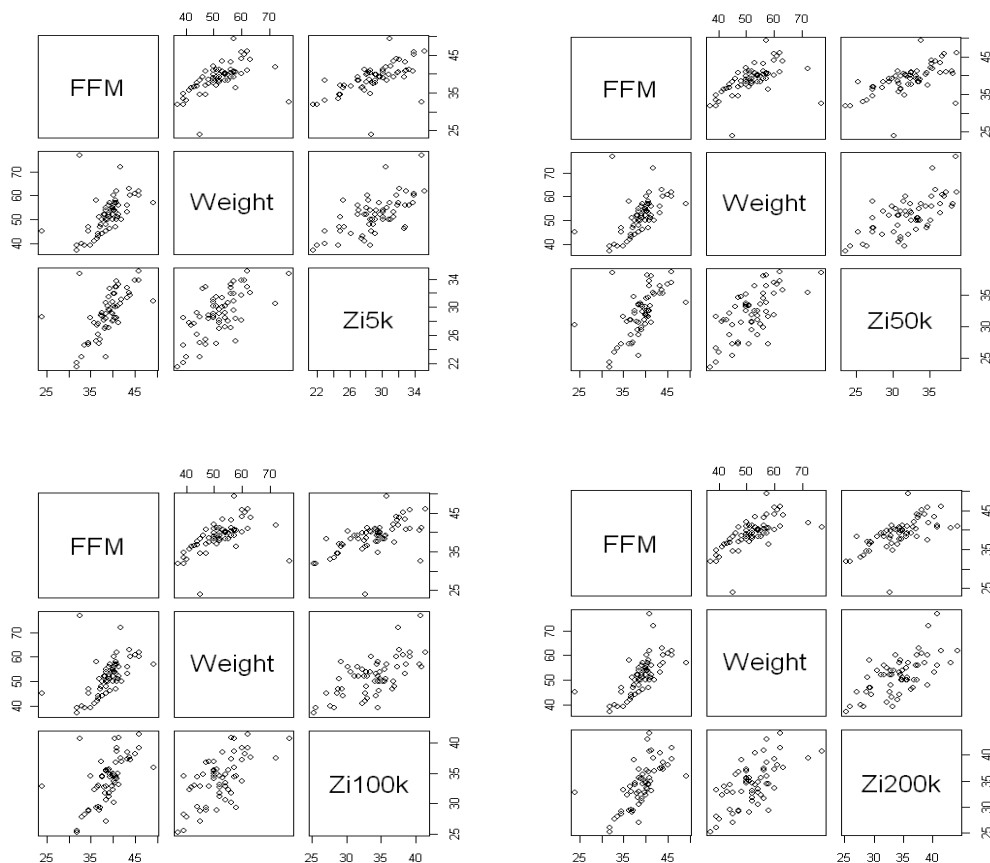
**Fig5:** Scale location plot between the square root of standardized residuals versus fitted values of Indian females showing the relationship between Total Body Water(TBW)in (litre) , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at 5KHz,50KHz.,100KHz.,200KHz.frequencies)in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg.



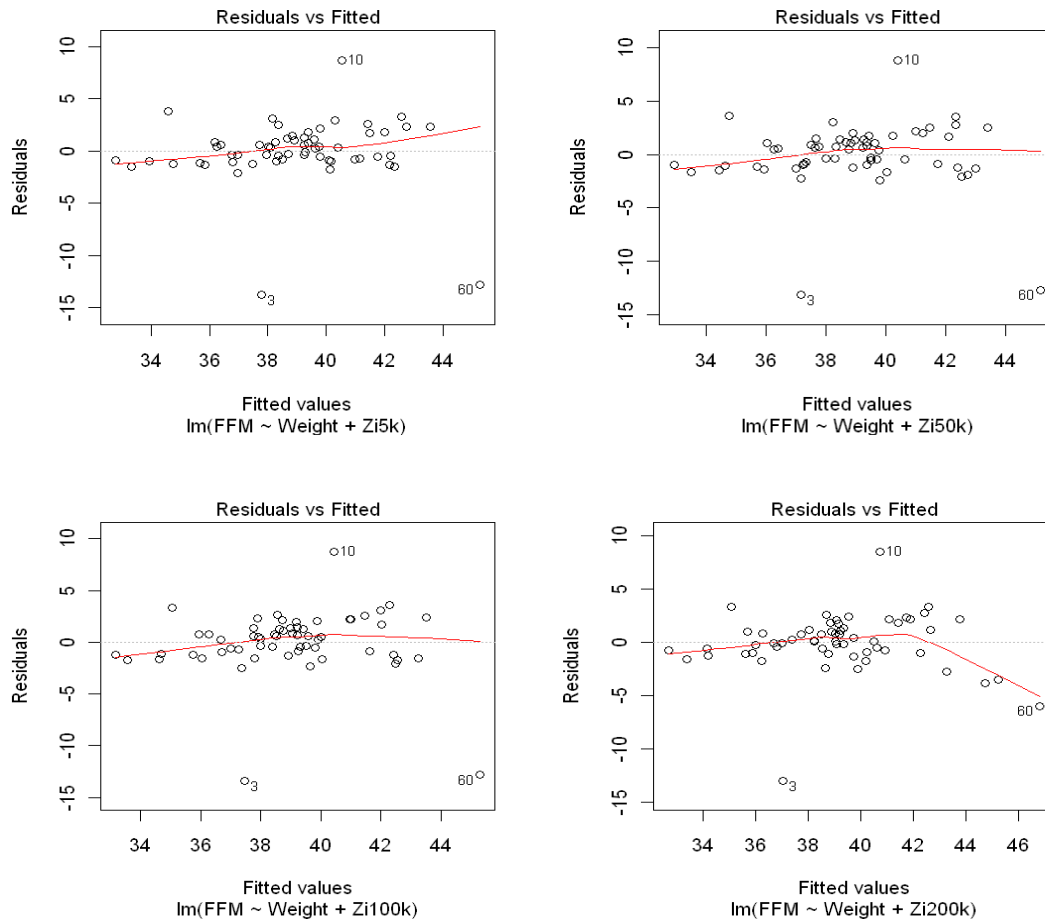




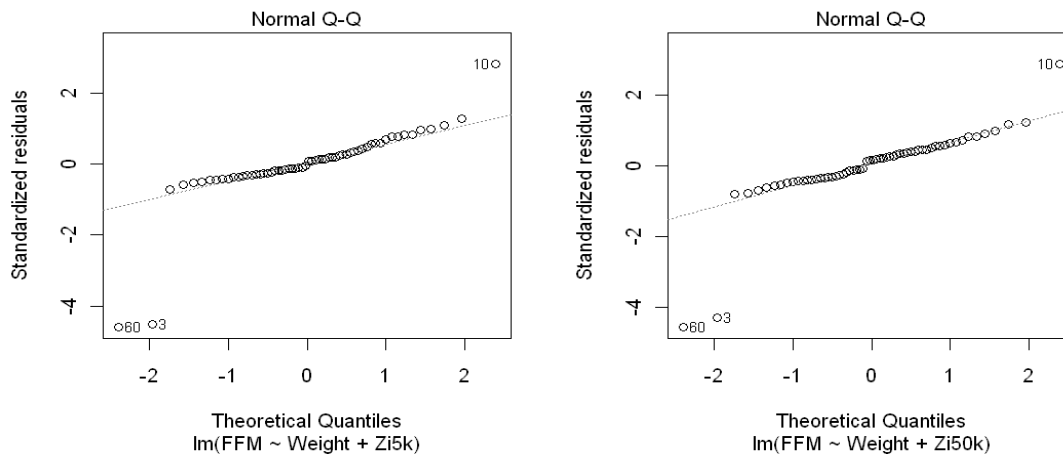
**Fig6:** Residual versus leverage plot and standardised residuals versus cook's distance plot of Indian females showing the relationship between Total Body Water(TBW)in (litre) , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at 5KHz,50KHz,100KHz,200KHz.frequencies)in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg.

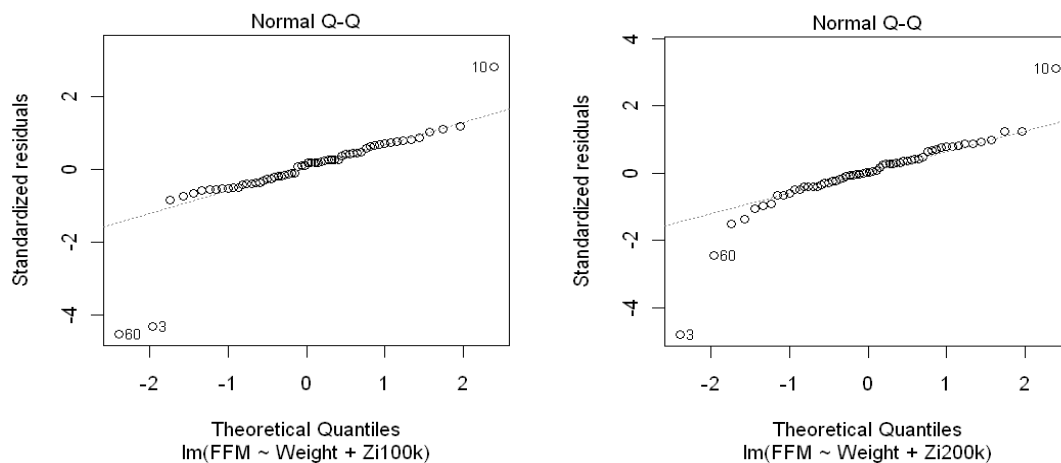


**Fig7:** Scatter Plot Matrix distribution of body composition of Indian females showing the relationship between Fat Free Mass(FFM)in Kg. , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at 5KHz,50KHz,100KHz,200KHzfrequencies) in ( $\text{cm}^2/\Omega$ ) and Weight of body.

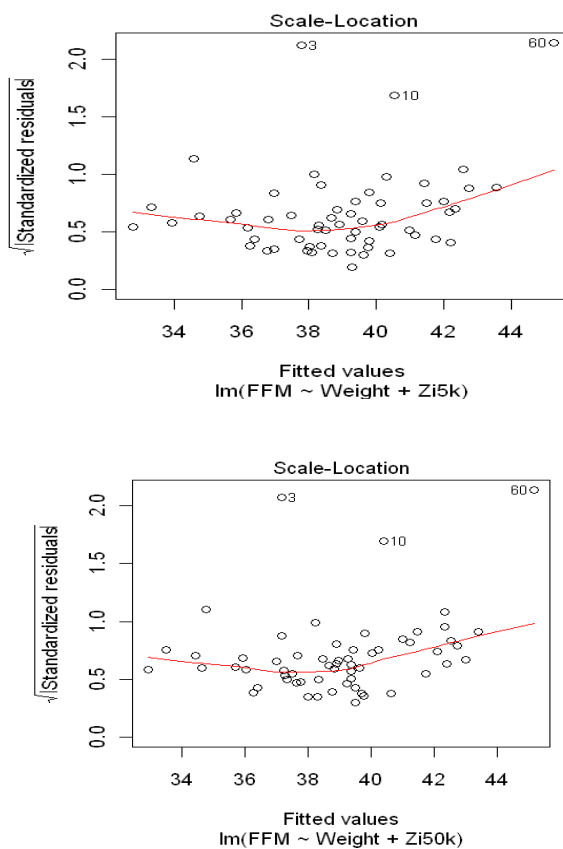


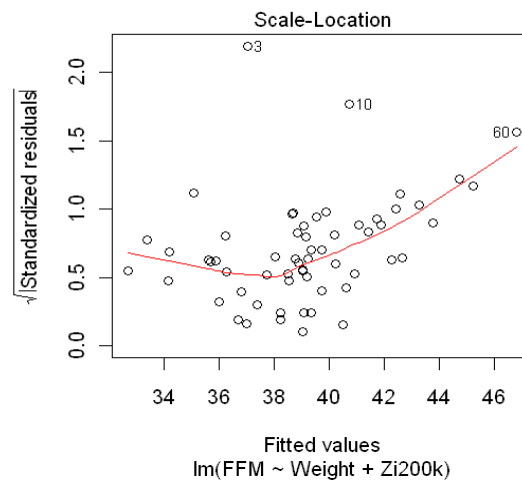
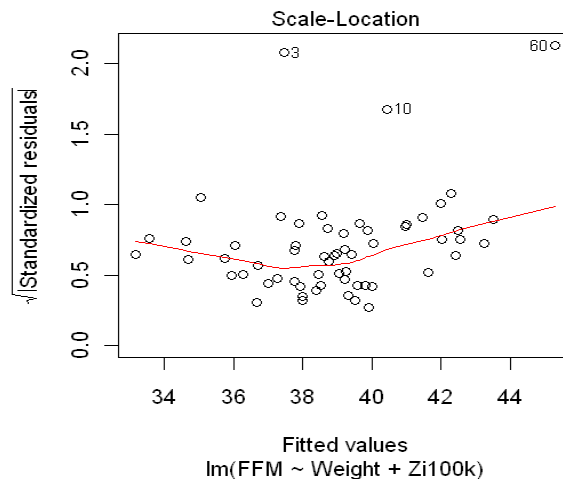
**Fig 8:** Random scatter distribution residual versus fitted values of Indian males showing the relationship between Fat Free Mass(FFM)in Kg. , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at 5KHz,50KHz,100KHz,200KHz.frequencies) in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg.



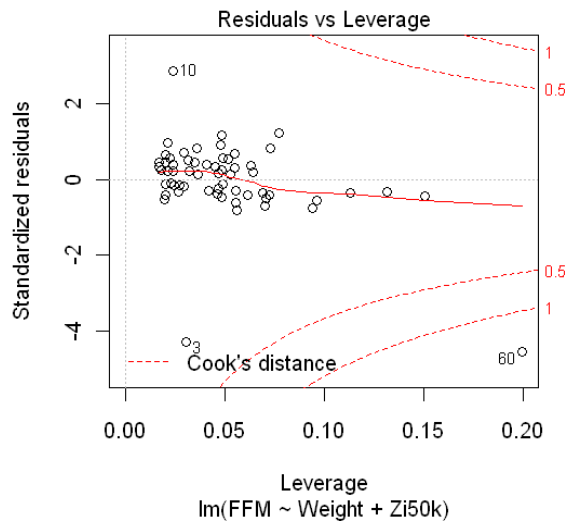
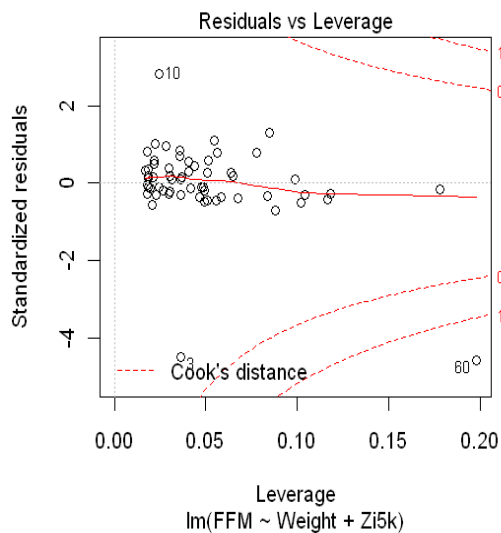


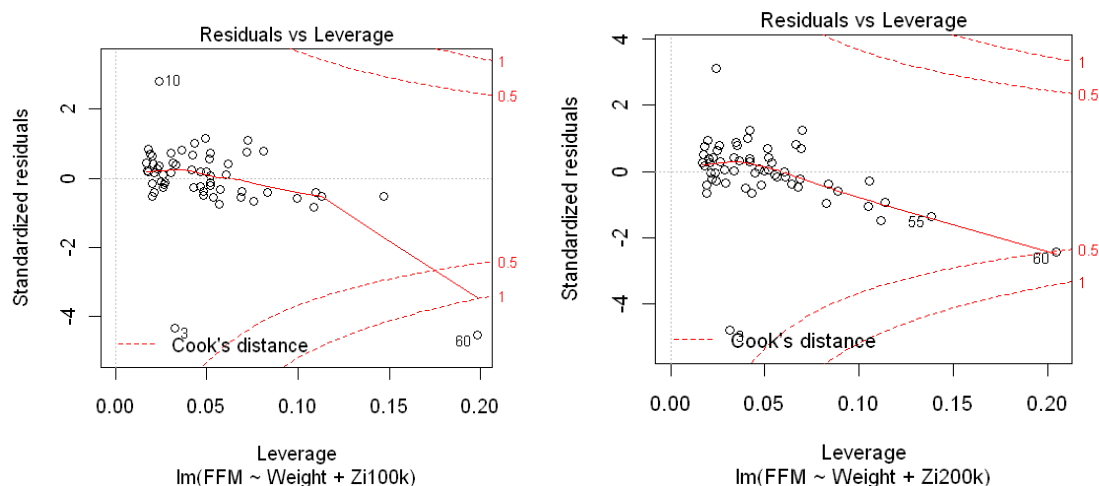
**Fig9:** Normal distribution versus Standardized residuals of Indian males showing the relationship between Fat Free Mass (FFM) in Kg., Impedance Index ( $\text{Height}^2/\text{Impedance of body}$  at 5KHz, 50KHz, 100KHz, 200KHz frequencies) in  $(\text{cm}^2/\Omega)$  and Weight of body in Kg.





**Fig10:** Scale location plot between the square root of standardized residuals versus fitted values of Indian females showing the relationship between Fat Free Mass(FFM)in (Kg) , Impedance Index ( $\text{Height}^2/\text{Impedance}$  of body at 5KHz,50KHz.,100KHz.,200KHz.frequencies)in ( $\text{cm}^2/\Omega$ ) and Weight of body in Kg.





**Fig11:** Residual versus leverage plot and standardised residuals versus cook’s distance plot of Indian females showing the relationship between Fat Free Mass(FFM)in (Kg) , Impedance Index (Height<sup>2</sup>/Impedance of body at 5KHz,50KHz,100KHz,200KHz.frequencies)in (cm<sup>2</sup>/Ω) and Weight of body in Kg.

**Comparison of measured and predicted results:** The Table below shows the comparative of measured and predicted values of TBW and FFM at frequencies of 5KHz,50KHz,100KHz and 200 KHz.

**TABLE 4:**

S.No.	TBW predicted at 5KHz.	TBW predicted at 50KHz	TBW predicted at 100 KHz	TBW predicted at 200 KHz	TBW measured	FFM predicted at 5KHz	FFM predicted at 50 KHz	FFM predicted at 100 KHz	FFM predicted at 200 KHz	FFM measured
1.	23.911	24.089	24.069	24.008	24.35	37.958	37.98	37.986	37.36	37.61
2.	25.38	25.82	25.736	26.132	26.22	38.688	39.365	39.247	39.72	38.38
3.	24.028	23.768	23.927	23.86	11.82	37.79	37.208	37.455	37.04	24.02
4.	26.687	27.373	27.209	27.557	27.79	40.308	41.24	41.005	41.418	43.27
5.	25.294	24.853	24.649	24.95	25.28	40.176	38.769	38.448	38.487	39.23
6.	30.497	30.028	30.025	29.917	30.19	43.57	43.41	43.503	43.76	45.91
7.	25.866	25.816	26.031	25.93	35.81	38.913	38.83	39.21	39.202	39.9
8.	22.505	22.830	22.928	22.58	23.27	36.95	36.98	37.37	35.89	34.87
9.	25.328	25.116	25.342	25.24	25.39	37.722	37.63	37.99	38.211	38.31
10.	27.553	27.5	27.529	27.412	27.38	40.539	40.39	40.45	40.744	49.24
11.	23.143	23.079	22.903	22.838	23.06	36.256	36.25	35.92	35.687	36.7
12.	26.212	26.039	26.051	25.96	26.04	40.101	39.427	39.52	39.355	39.2
13.	21.516	21.349	21.401	21.357	21.64	40.432	34.608	34.677	34.142	33.54
14.	23.275	23.558	23.511	23.452	24.03	36.969	37.34	37.279	26.227	36.59
15.	26.405	26.033	26.086	25.98	25.73	39.263	38.675	38.745	39.035	39.86
16.	25.878	26.321	26.102	26.016	26.24	39.784	40.225	39.89	39.55	41.94
17.	27.154	27.053	27.443	27.32	27.16	39.235	39.36	40.02	40.496	40.56
18.	21.329	21.394	21.518	21.47	21.6	33.925	34.42	34.603	34.011	32.93
19.	25.87	25.338	25.534	25.426	25.22	38.078	37.484	36.552	38.238	38.4
20.	28.286	28.268	28.178	28.083	28.66	42.163	41.713	41.63	41.71	44.04
21.	26.333	26.07	26.179	26.07	25.85	38.68	38.339	38.629	39.045	39.88
22.	24.529	24.244	24.428	24.33	24.21	36.373	36.4	36.66	37.017	36.95
23.	26.304	25.547	25.443	25.35	25.93	38.36	38.89	38.75	39.15	40.89
24.	27.621	27.688	27.660	27.55	27.51	41.477	40.98	40.96	41.055	43.19
25.	29.346	29.42	29.369	29.227	29.13	41.98	42.104	42.018	42.65	43.77
26.	24.709	24.57	24.67	24.59	24.35	38.04	37.768	37.928	37.74	38.47
27.	27.221	27.089	27.167	27.05	27.25	39.804	39.73	39.812	40.22	39.25
28.	23.07	22.9	22.947	22.87	23.13	35.673	35.68	35.73	35.627	34.56
29.	25.798	25.73	25.625	25.538	26.08	39.604	39.218	39.05	38.87	39.88
30.	27.92	28.51	28.49	28.381	28.81	41.777	42.39	42.43	42.25	41.2
31.	27.596	27.585	27.696	27.565	27.62	39.370	39.75	39.907	40.613	40.13
32.	25.882	26.003	25.866	25.774	26.24	39.371	39.415	39.197	39.09	41.18
33.	26.622	26.43	26.563	26.458	26.62	39.74	39.36	39.586	39.72	40.16
34.	29.049	29.09	29.045	28.917	28.9	42.589	42.33	42.285	42.55	45.89
35.	28.522	28.583	28.574	28.93	28.3	41.42	41.46	41.455	41.877	44
36.	29.314	29.247	29.029	28.89	29.01	42.75	42.336	41.98	42.41	45.1

37.	25.53	25.469	25.667	25.57	25.35	38.17	41.238	38.566	38.679	41.25
38.	26.033	25.131	25.215	25.116	25s	39.28	37.656	37.76	38.03	39.17
39.	26.329	26.133	26.27	26.155	26.32	38.25	38.319	38.51	39.05	39.08
40.	23.684	23.403	23.573	23.496	23.51	36.19	36.016	36.27	36.28	37.06
41.	20.236	20.04	20.217	20.18	20.75	32.77	32.91	33.149	32.68	31.91
42.	25.613	25.5	25.563	25.465	25.85	38.35	38.29	38.38	38.53	37.91
43.	25.589	25.51	25.361	25.29	25.49	40.41	39.626	39.42	38.856	40.7
44.	26.36	26.22	26.201	26.1	25.99	39.59	39.27	39.23	39.325	40.67
45.	20.867	20.696	20.759	20.825	21.06	33.325	33.488	33.55	33.38	31.81
46.	25.388	25.547	25.557	25.47	25.74	38.48	38.898	38.93	38.77	37.66
47.	26.358	25.807	26.175	26.05	25.7	37.47	37.232	37.78	38.67	36.27
48.	22.823	23.048	22.86	22.8	23.07	36.814	36.99	36.69	35.98	35.71
49.	22.984	23.355	23.196	23.47	23.88	36.748	37.25	37.003	36.822	36.4
50.	23.032	22.35	22.559	22.49	25.96	34.574	34.75	35.042	35.07	38.4
51.	23.526	23.347	23.448	23.47	23.5	35.84	35.919	36.05	36.25	34.51
52.	26.216	26.208	26.227	26.12	26.21	38.84	38.975	38.99	39.24	40.34
53.	29.108	29.63	29.521	29.85	29.6	42.33	42.73	42.56	43.67	40.84
54.	24.825	25.75	25.644	26.08	26.25	38.29	39.78	39.644	39.88	37.34
55.	31.159	31.33	31.497	31.75	31.27	42.2	43.002	43.25	45.22	41.73
56.	26.99	26.975	26.98	26.87	25.96	40.138	40.024	40.04	40.199	38.4
57.	27.832	28.754	28.68	29.155	29.11	41.159	42.546	42.49	43.25	40.47
58.	27.568	27.472	25.91	27.466	27.61	40.99	40.614	37.88	40.94	40.18
59.	25.639	25.893	28.156	25.69	26.02	39.25	39.499	39.326	39.09	38.93
60.	33.504	33.371	33.45	33.26	32.48	45.178	45.178	45.316	46.81	40.8

**Results and Discussion:** The physical characteristics of samples are shown in Table 1 and Table 2. The prediction equations developed for Total Body Water at 5 KHz, 50 KHz, 100 KHz and 200 KHz. are as shown below:

$$TBW_{BIA} = (0.24689) \times ZI_{5k} + (0.25593) \times \text{Body weight} + 5.58336 \dots\dots\dots(1)$$

$$TBW_{BIA} = (0.0.22363) \times ZI_{50k} + (0.29178) \times \text{Body weight} + 4.89741 \dots\dots\dots(2)$$

$$TBW_{BIA} = (0.23751) \times ZI_{100k} + (0.24204) \times \text{Body weight} + 5.31032 \dots\dots\dots(3)$$

$$TBW_{BIA} = (0.0.22363) \times ZI_{200k} + (0.29178) \times \text{Body weight} + 4.89741 \dots\dots\dots(4)$$

Where  $TBW_{BIA}$  is Total body water in litres,  $ZI_{5k}$ ,  $ZI_{50k}$ ,  $ZI_{100k}$ ,  $ZI_{200k}$  is Impedance indexes of the body at 5KHz, 50 KHz, 100KHz, 200 KHz respectively in ( $cm^2/\Omega$ ). Body weight is the weight of the body in Kg. The prediction equations developed for Fat Free Mass at 5 KHz, 50 KHz, 100 KHz and 200 KHz. are as shown below:

$$FFM_{BIA} = (0.12496) \times ZI_{5k} + (0.56707) \times \text{Body weight} + 15.91914 \dots\dots\dots(5)$$

$$FFM_{BIA} = (0.11924) \times ZI_{50k} + (0.50014) \times \text{Body weight} + 16.70689 \dots\dots\dots(6)$$

$$FFM_{BIA} = (0.13920) \times ZI_{100k} + (0.42686) \times \text{Body weight} + 17.20722 \dots\dots\dots(7)$$

$$FFM_{BIA} = (0.21844) \times ZI_{200k} + (0.34863) \times \text{Body weight} + 15.78796 \dots\dots\dots(8)$$

Where  $FFM_{BIA}$  is Total body water in Kg,  $ZI_{5k}$ ,  $ZI_{50k}$ ,  $ZI_{100k}$ ,  $ZI_{200k}$  is Impedance indexes of the body at 5KHz, 50 KHz, 100KHz, 200 KHz respectively in ( $cm^2/\Omega$ ). Body weight is the weight of the body in Kg. The other statistical analysis such as Standard Error of Estimate (S.E.E.) for Intercept, Impedance index at 5KHz, 50KHz, 100KHz and 200 KHz, Residual Standard Error (R.S.E.), Multiple R squared, mean standard deviation of predictor variable i.e. Impedance index and weight and dependent variable i.e. TBW and FFM at 5KHz, 50KHz, 100KHz and 200 KHz are given in Table 3.

A comparative study of TBW and FFM at frequencies of 5 KHz, 50 KHz, 100 KHz and 200 KHz is shown in Table 4. From the results obtained it is seen that predicted values are very close to measured values. However, predicted values of TBW and FFM at 50 KHz, 100 KHz and 200 KHz are much closer to measured values than the predicted values at 5 KHz. This is due to the fact that at low frequencies, the current cannot bridge the cellular membrane and will pass predominantly through the extracellular space. At higher frequencies penetration of the cell membrane occurs and the current is conducted by both the extra-cellular water (ECW) and intra-cellular water (ICW).

#### REFERANCES

- [1] P.N. Goswami Thesis under the supervision of Dr. Munna Khan ;Prediction and measurement of Human Body composition using Non-Invasive technique.
- [2] Segal KR, Van Loan M, Fitzgerald PI, Hodgdon JA, Van Itallie TB. Lean body mass estimation by bioelectrical impedance analysis:a four site cross validation study. *Am J Clin Nutr* 1988;47:7-14
- [3] Kushner RF, Schoeller DA. Estimation of body water by bioelectrical impedance analysis. *Am J Clin Nutr* 1986;44:417-24
- [4] Deurenberg P, Van der Kooy K, Leenen R, Weststrate JA, Seidell JC. Sex and age specific prediction formulas for estimating body composition from bioelectrical impedance: a cross-validation study. *Int J Obes* 1991;15:17-25
- [5] Jean C Desport, Pierre M Preux, Corinne Boutloup-Demange, Pierre Clavelou, Bernard Beaufriere, Christine Bonnet, and Philippe P Couratier. Validation of bioelectrical impedance analysis in patients with amyotrophic lateral sclerosis<sup>1-3</sup>. *The American Journal of Clinical Nutrition* (2003).
- [6] Verena K Hass, Jane R Allen, Michael R Kohn, Simon D Clark, ShuHua Zhang, Julie N Briody, Margie Gruca, Sloane Madden, Manfred J Muller, and Kevin J Gaskin. Total body protein in healthy adolescent girls: validation of estimates derived from simpler measures with neutron activation analysis<sup>1-3</sup>. *The American Journal of Clinical Nutrition* (2007).
- [7] P Deurenberg, M Deurenberg-Yap, LF Foo, G Schmidt and J Wang. Differences in body composition between Singapore Chinese, Beijing Chinese and Dutch children. *European Journal of Clinical Nutrition* (2003).
- [8] Jordan R Moon, Holly R Hull, Sarah E Tobkin, Masaru Teramoto, Murat Karabulut1, Michael D Roberts, Eric D Ryan, So Jung Kim, Vincent J Dalbo, Ashley A Walter, Abbie T Smith, Joel T Cramer and Jeffrey R Stout. Percent body fat estimations in college women using field and laboratory methods: a three-compartment model approach. *Journal of the International Society of Sports Nutrition* 2007, doi:10.1186/1550-2783, pp:4-16
- [9] Noe'l Cameron, Paula L Griffiths, Melanie M Wright, Charlotte Blencowe, Nicola C Davis, John M Pettifor, and Shane A Norris. Regression equations to estimate percentage body fat in African prepubertal children aged 9. *Am J Clin Nutr* 2004, Vol. 80, No. 1, pp: 70-75, July 2004
- [10] Cameron B Ritchie and Robert T Davidson\* Regional body composition in college-aged Caucasians from anthropometric measures, *Nutrition & Metabolism* 2007, doi:10.1186/1743-7075, pp:4-29
- [11] Carlo Basile, Luigi Vernaglione, Biagio Di Iorio, Vincenzo Bellizzi, Domenico Chimienti, Carlo Lomonte, Anna Rubino, and Nicola D'Ambrosio. Development and Validation of Bioimpedance Analysis Prediction Equations for Dry Weight in Hemodialysis Patients. *Clin J Am Soc Nephrol* 2: pp:675-680, 2007
- [12] Amir Haider Shah and Rakhshanda Bilal; Body composition, its Significance and model for assessment: *Pakistan Journal of Nutrition* 2009. Vol-2, pp:198-202
- [13] Nyboer J. Percent body fat by four terminal bio-electrical impedance and body density in college freshmen. *Proceedings of the V International Conference on Electrical Bio-impedance* 1981, Tokyo., Vol-8, pp:56-71
- [14] Sun SS, Chumlea WC, Heymsfield SB, Lukaski HC, Schoeller D, Friedl K, Kuczmarski RJ, Flegal KM, Johnson CL & Hubbard Van S (2003) Development of bioelectrical impedance analysis prediction equations for body composition with the use of a multicomponent model for use in epidemiological surveys, Vol-77, pp:331-340
- [15] Troiano RP, Flegal KM, Kuczmarski RJ. Overweight prevalence and trends for children and adolescents: The National Health and Nutrition Examination Surveys, 1963 to 1991. *Arch Pediatr Adolesc Med* 1995; Vol-149:1085-91

**Author for Corrospendence:** Ghazala Perveen Khan,

# The Applications of Computational Intelligence (Ci) Techniques in System Identification and Digital Filter Design

Jainarayan Yadav<sup>1</sup>, Sanjay Gurjar<sup>2</sup>

<sup>1</sup>M.Tech (Ece), Bhagwant University Ajmer & Mb No: 07542029216,

<sup>2</sup>Asst.Prof. (Ece), Bhagwant University Ajmer & 07597730006

## ABSTRACT

*The thesis focuses on the application of computational intelligence (CI) techniques for two problems- System identification and digital filter design. In system identification, different case studies have been carried out with equal or reduced number of orders as the original system and also in identifying a blackbox model. Lowpass, Highpass, Bandpass, stopband FIR and Lowpass IIR filters have been designed using three algorithms. Using two different fitness functions. Particle Swarm Optimization (PSO), Differential Evolution based PSO (DEPSO) and PSO with Quantum Infusion (PSO-QI) algorithms have been applied in this work. PSO-QI is a new Hybrid algorithm where the global best particle obtained from PSO goes into a tournament with an offspring produced by mutating the global best of PSO using the quantum principle in quantum behaved PSO (QPSO) and the winner is selected as the new global best of the swarm.*

*In QPSO, unlike traditional PSO, exact values of particle's position and velocity cannot be determined. However, its position in the solution space is determined by mapping the probability of its appearance in the quantized search space. The results obtained from PSO-QI have been compared with the DEPSO hybrid algorithm and the classical PSO. In all of the cases, PSO-QI has outperformed the other two algorithms in its ability to converge to the lowest error value and its consistency in finding the solution every time and thus proven to be the best. However, the computational complexity of PSO-QI is higher than that of the other two algorithms.*

## I. INTRODUCTION

System identification is a challenging and complex optimization problem due to nonlinearity of systems and even more in a dynamic environment. Adaptive infinite impulse response (IIR) systems are preferably used in modeling real world systems because of their reduced number of coefficients and better response over the finite impulse response (FIR) filters. In this work, system identification has been viewed as a problem of adaptive IIR filtering so that it becomes a parameter estimation problem.

Digital filter design is also a complex optimization problem due to the number of filter parameters that can be optimized. Hence CI techniques can be used to estimate the filter coefficients so as to optimize these parameters and design the desired filter response.

PSO and its other variants have been a topic of research over the past decade. Inspired by social behavior of bird flocking and fish schooling, PSO has proven to be an effective stochastic search technique. Hence it has been applied to a wide variety of problems related to search optimization, clustering, routing, scheduling. PSO has gone through various changes and different variants have been introduced in order to solve the problem more effectively. It has also been combined with other different algorithms to create hybrid optimization algorithms. These algorithms have been reported in different literatures and applied to different practical applications. In this thesis, two problems have been studied – system identification and digital filter design. These applications have been implemented using the standard PSO and two hybrid algorithms – differential evolution particle swarm optimization (DEPSO) and PSO with quantum infusion (PSO-QI). These results of system identification have also been compared with another hybrid algorithm PSO with evolutionary algorithm (PSO-EA). The thesis covers the details of these algorithms, the research work carried out towards the implementation of the above mentioned problems and their results.



## **II. OBJECTIVES**

The main objective of this research is to apply swarm, evolutionary and quantum based algorithms to solve two practical problems viz. system identification and digital filter design. PSO, DEPSO and PSO-QI are the major algorithms involved in this work for system identification and in the design of digital filters. The results of the case studies are also presented

## **III. THESIS LAYOUT**

The thesis has been divided into 9 chapters. Chapter 1 introduces to the topic and Major areas of this research work. In Chapter 2, system identification has been explained This chapter introduces to the problem of system identification and traditional and modern techniques used to solve it. In Chapter 3, digital filter design is explained. This chapter introduces to the problem and traditional and modern techniques used in digital filter design In next three chapters, the three algorithms have been explained in detail. In Chapter 4, PSO has been covered. This chapter explains the basics of the algorithm and how it has been applied to the above mentioned problems. In Chapter 5, DEPSO has been explained. Similarly, PSO-QI has been explained in Chapter 6 In the next two chapters, case studies carried out during the research and the results obtained from them have been presented. In Chapter 7, studies and results of system identification have been presented. This chapter shows the comparison of results obtained from system identification, and is presented as figures and tabulated data. In Chapter 8, similar results obtained for digital filter design are presented. These results are also presented as figures and tabulated data and show a comparison of different algorithms as applied to the problem. Conclusion of the thesis and future work is presented

## **SYSTEM IDENTIFICATION**

### **INTRODUCTION**

System identification is a challenging and complex optimization problem due to nonlinearity of the systems and even more in a dynamic environment. Adaptive infinite impulse response systems are preferably used in modeling real world systems because of their reduced number of coefficients and better performance over the finite impulse response filters. Particle Swarm Optimization (PSO) and its other variants has been a subject of research for the past few decades for solving complex optimization problems. In this thesis, the concept of Differential Evolution based Particle Swarm Optimization (DEPSO) is implemented for system identification. A hybrid of Particle Swarm Optimization and Evolutionary Algorithm (PSO-EA) has been considered for comparison with PSO and DEPSO algorithms.

### **SYSTEM IDENTIFICATION PROBLEM**

System identification is the mathematical modeling of an unknown system by monitoring its input output data. This is achieved by varying the parameters of the enveloped model so that for a set of given inputs, its output match that of the system under consideration. For a plant whose behavior is not known, an adaptive system can be modeled and its parameters can be continuously adjusted using any adaptive algorithms. By the use of such adaptive algorithms, the required parameters can be obtained such that the output of the plant and the model are same for the same set of inputs, which is the goal of system identification (Panda et al., 2007). Traditionally, Least Mean Square(LMS) and other algorithms have been studied for the identification of linear and static systems (Windrow et. al., 1976). But, almost all physical systems are nonlinear to certain extent and recursive in nature and hence it is more convincing to model such systems by using nonlinear models (Panda et. al., 2007; Krusienski and Jenkins, 2005). Thus nonlinear system identification has attracted attention in the field of science and engineering. Hence these are better modeled as Infinite Impulse Response (IIR) models as they can provide better performance than a Finite Impulse Response (FIR) filter with the same number of coefficients (Shynk, 1989(a)). Thus the problem of nonlinear system identification can also be viewed as a problem of adaptive IIR filtering. Also, IIR models are more efficient than the FIR models for implementation as they require less parameter and hence fewer computations for the same level of performance. However, there are few problems associated with the use of IIR models in identification of a system, such as instability of the algorithms, slow convergence and convergence to the local minimum(Netto et al., 1995). In order to overcome these, different techniques have been developed over the years. Fig. shows a block diagram describing the problem of system identification.

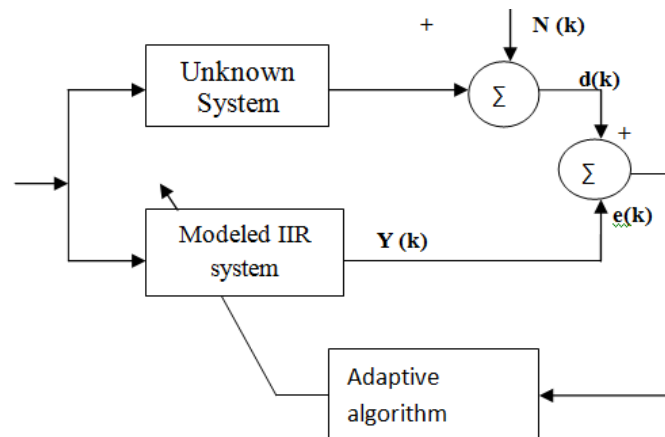


Figure.: Schematic showing system identification

### SYSTEM IDENTIFICATION TECHNIQUES

Different learning algorithms have been used in the past for nonlinear system identification. These techniques include use of neural network (Hongwei and Yanchun, 2005) and gradient based search techniques such as least mean square algorithm (Shynk, 1989(a)). Unfortunately, the error surface of such recursive systems such as a multi-machine power system (Kundur, 1993) tends to be multi-modal and hence traditional techniques of parameter approximation fail as they get trapped into local minimum and cannot attain the global minimum (Krusienski and Jenkins, 2005). Various algorithms that are implemented in the adaptive IIR filtering for system identification are described in (Netto et al., 1995).

Population based search algorithm such as Genetic Algorithm (GA) has also been used for the system identification. It uses a population of potential solutions encoded as chromosomes which go through genetic operations such as crossover and mutation to find the best solution (Kristinsson and Dumont, 1992). But its effectiveness is affected by the convergence time (the time it takes to find the global minimum). So to eliminate such inefficiencies, population based stochastic optimization techniques have been discussed in various literatures. Particle Swarm Optimization (PSO) is one of the most known techniques (delValle et al., 2007). Application of PSO in the system identification has been discussed in (Panda et al., 2007). In (Lee et al., 2006), a method for the identification of nonlinear system and parameter optimization of the obtained input-output model has been described. The proposed method uses least squares support vector machines regression based on PSO. In another work, PSO has been used for optimizing the parameters of Elman neural network which is used for speed identification of ultrasonic motors (Hongwei and Yanchun, 2005). A modified form of PSO called as the self-organizing particle swarm optimization and its application in the system identification has been discussed in (Shen and Zeng, 2007). Radial Basis Function Neural Network (RBFNN) has been used for system identification in (Chen et al., 2007), where a hybrid gradient-based PSO algorithm has been used to adjust the parameters of the RBFNN. In (Liu et al., 2006), particle swarm optimization and quantum-behaved particle swarm optimization have been used for the system identification. Use of different types of stochastic optimization techniques in adaptive IIR filters and nonlinear systems has been explained in (Krusienski and Jenkins, 2005). Use of Differential Evolution (DE) and Ant Colony Optimization (ACO) in IIR filter design has been presented in (Karaboga, 2005) and (Karaboga et al., 2004) respectively. They also talk about the possible use of these approaches in system identification and other applications. But these algorithms have the tendency to get stuck in the local minimum when the complexity of the problem increases and in dynamic systems where time allowed for convergence is constrained. Hybrid algorithms are used to improve the performance by combining the best feature of both algorithms. In (Cai et al., 2007), one such hybrid algorithm has been shown. In the paper, PSO and Evolutionary Algorithm (PSO-EA) hybrid has been implemented to combine the best features of PSO (co-operation) and EA (competition).

### SUMMARY

Identification of complex systems is an optimization problem and is viewed as IIR system identification in this chapter. By the use of swarm and evolutionary algorithms, the coefficients of the filter are determined. The results of the study are shown in Chapter

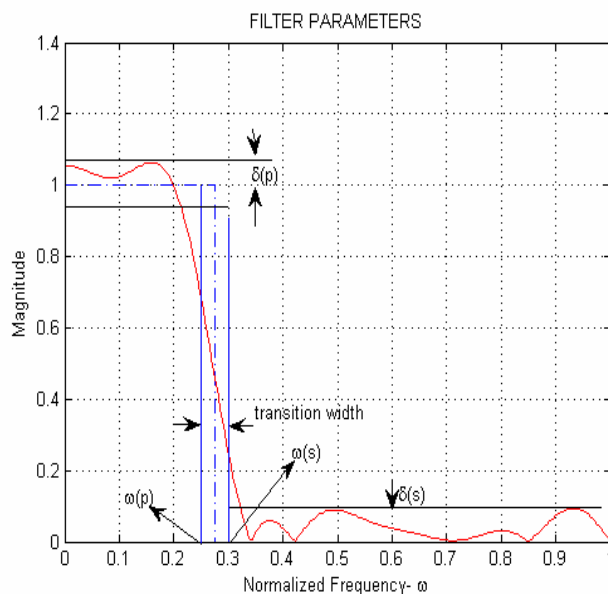
## DIGITAL FILTER DESIGN

### INTRODUCTION

This chapter introduces digital filter design as an optimization problem and discusses various methods applied in the design of digital filters traditionally and currently using the computational intelligence techniques.

### DIGITAL FILTER

A filter is a frequency selective circuit that allows a certain frequency to pass while attenuating the others. Filters could be analog or digital. Analog filters use electronic components such as resistor, capacitor, transistor etc. to perform the filtering operations. These are mostly used in communication for noise reduction, video/audio signal enhancement etc. In contrast, digital filters use digital processors which perform mathematical calculations on the sampled values of the signal in order to perform the filter operation. A computer or a dedicated digital signal processor may be used for implementing digital filters. Filters mostly find their use in communication for noise reduction, audio/video signal enhancement etc. Any time varying signal  $C=x(t)$  sampled at a sampling interval of  $h$  has input signals  $X_0, X_1, X_2, \dots$  in intervals  $0, h, 2h, 3h, \dots, nh$ . These inputs have corresponding outputs  $y_0, y_1, y_2, \dots, y_n$  depending upon the kind of operation performed. Thus, the order of the filter is determined by the number of the previous input terms used to calculate the current output. The  $a_0, a_1, a_2$  terms appearing in the following equations are called the filter coefficients and determine the operation of the filter. These determine the characteristics of the filter. Various filter parameters which come into picture are the stopband and passband normalized frequencies ( $\omega_s, \omega_p$ ) the passband and stopband ripple ( $\delta_p$ ) and ( $\delta_s$ ) the stopband attenuation and the transition width. This has been shown in Fig.



## PARTICLE SWARM OPTIMIZATION

### INTRODUCTION

Introduced by Eberhart and Kennedy in 1995 (del Valle et al., 2007), PSO is a search technique based on social behavior of bird flocking and fish schooling. There are different kinds of bio and social behavior inspired algorithms. PSO is one of the different swarm based algorithms. In PSO, each particle of the swarm is a possible solution in the multi-dimensional search space. The particles change their positions with a certain velocity in each iteration, according to the standard PSO equations, thus moving towards the global best (gbest) solution. Being easy to implement and yet so effective, PSO has been utilized in a wide variety of optimization applications. In this thesis, PSO has been used in system identification and to design digital filters.

### PSO ALGORITHM

Particle swarm optimization is a population based search algorithm and is inspired by the observation of natural habits of bird flocking and fish schooling. In PSO, a swarm of particles moves through a D dimensional search space. The particles in the search process are the potential solutions, which move around the defined search space with some velocity until the error is minimized or the solution is reached, as decided by the fitness function. Fitness function is the measure of particles fitness which is the deviation of the particle from the required solution. The particles reach to the desired solution by updating their position and velocity according to the PSO equations. In PSO model, each individual is treated as a volume-less particle in the D - dimensional search space with initial random velocity. Each particle has memory which keeps track of its previous best position and fitness, with the position and velocity of i particle represented as:

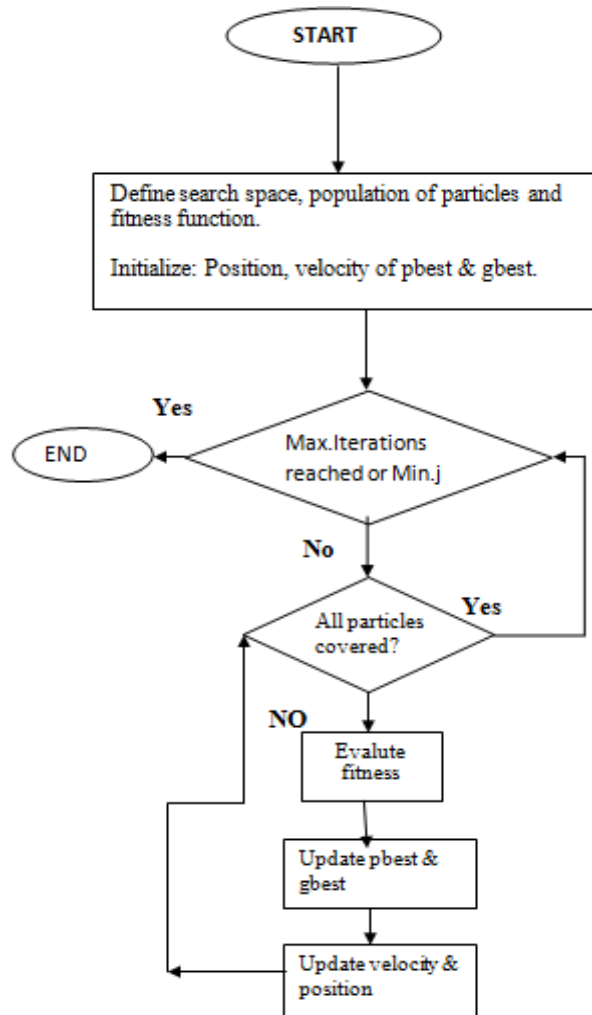


Fig: Flowchart for PSO

### DIFFERENTIAL EVOLUTION PARTICLE SWARM OPTIMIZATION INTRODUCTION

Due to the limitations of PSO in finding the best solution, different other approaches were also considered. Over the past few years, research in the field of computational intelligence gave birth to a number of different approaches. All of these algorithms had some special features in finding the best solutions, either their convergence speed or their ability to find the better solution. However, they suffered from one or the other problems. In order to overcome these shortcomings and utilize their effective best properties, hybrid algorithms were introduced. Hybrid algorithms take the best features of the individual algorithms and thus tend to be more effective than the individual algorithms. DEPSO is one of such hybrid algorithms. In this chapter, DEPSO and its applications in system identification and digital filter design is discussed.

## **DEPSO ALGORITHM**

DEPSO is the hybrid of DE and PSO. Differential Evolution. Differential Evolution was introduced by Storn and Price in 1995 (Storn, 1996). It is also a population based stochastic search technique for function minimization. In (Storn, 1996), DE has been applied in the field of filter design. In DE, the weighted difference between the two population vectors is added to a third vector and optimized using selection, crossover and mutation operators as in to a third vector and optimized using selection, crossover and mutation operators as in individual, called the offspring, is then recombined with the parent under certain criteria such as crossover rate. Fitness of both the parent and the offspring is then calculated and the offspring is selected for the next generation only if it has a better fitness than the parent (Karaboga, 2005). The mutation takes place according to (14).

## **CONCLUSION INTRODUCTION**

In this work, swarm, evolutionary and quantum based intelligent optimization algorithms are used in system identification and to design digital filters. It was shown that the swarm based algorithms has many variants and has been hybridized with other algorithms to increase its effectiveness. It was also seen that by hybridization of the algorithms, best features of both the algorithms are retained and thus new algorithm so developed is more robust. In this chapter, a conclusion of all the chapters is provided.

## **SECTION SUMMARY**

The first three chapters of the thesis cover the introduction to the problem. In Chapter 1, introduction to the thesis is provided. Chapter 2 covers the description of system identification. Introduction to the problem and traditional and modern methods applied to solve it are explained in this chapter. In Chapter 3, digital filter design is explained. This chapter also introduces to the problem of digital filter design and various traditional and new methods applied in the design. The next three chapters of the thesis describe the involved algorithms and their operation in detail. In Chapter 4, particle swarm optimization has been explained. As one of the pioneer stochastic search optimization technique based on the social behavior of bird flocking and fish schooling, algorithm of PSO has been described in this chapter. In Chapter 5, a hybrid optimization algorithm DEPSO has been explained. A combination of DE and PSO, it uses the differential evolution operation on the particle of the PSO to mutate the particle and create an offspring. The chapter covers the detail of its operation. In Chapter 6, another hybrid algorithm, PSO-QI has been introduced. PSO-QI emerges from the infusion of quantum operation obtained from QPSO on the particle of the PSO. Concepts of quantum particle swarm optimization and its application on the PSO have been explained in this chapter. In the next two chapters, the results obtained from the case studies have been presented. In Chapter 7, the results obtained from the application of different algorithms in the system identification have been presented. These results show the effectiveness of the new hybrid algorithms in comparison to the traditional PSO. In Chapter 8, the results for the digital filter design are shown. This chapter shows the results of designing different kinds of digital filters using various algorithms described in the previous chapters. These results also suggest that the new hybrid algorithms are more effective than the traditional PSO. These two chapters present their comparison in terms of figures and tabulated data from different case studies.

## **MAIN CONCLUSION**

The main focus of the thesis is in system identification and in the design of digital filters. The research work leading to the thesis is related to identification of an IIR system. This is achieved by modeling the unknown system with IIR systems of same or reduced number of orders. In digital filter design, Lowpass, Highpass, Bandpass and Bandstop FIR and Lowpass IIR filters are designed using different optimization algorithms. The results for system identification as well as digital filter design have been shown. In this work, particle swarm optimization is used as the baseline algorithm. Two other algorithms are considered to improve the results obtained from PSO. These are hybrid algorithms based on differential evolution and quantum particle. The DEPSO algorithm performed better than PSO in system identification as well as in digital filter design. Results obtained from PSO-QI are better than both PSO and DEPSO and hence it has outperformed the other two algorithms in all the case studies of system identification and digital filter design.

Fitness function based on passband and stopband ripples of the filter response is used to design both FIR and IIR filters whereas the fitness function based on MSE is used to design FIR filters only. It is observed that all three of the algorithms are able to approximate the filter coefficients in a number of iterations but PSO-QI always performed the best among them. Figures and tabulated results all show that PSO-QI is more consistent in its performance and it can achieve a lower value of average error in the cases using two different fitness functions. Although it took longer for the algorithm to converge because of its computational complexity, it found much better solution than PSO, DEPSO and QPSO. The results are not tabulated for QPSO because of the higher number of iterations and the results are clear from the figures. However, comparison has been made to confirm that PSO can not achieve the same amount of convergence even when allowed to run for the amount of time taken by PSO-QI. Hence, it can be concluded that swarm, evolutionary and quantum algorithms can be effectively used in digital filter design, and PSO-QI is a better choice. It is evident from the figures and results how the best features of two algorithms can be extracted and performance can be improved by the hybridization of these algorithms.

## **FUTURE RESEARCH**

This thesis covered application of different optimization algorithms in system identification and digital filter design problems. However, there is more room for research. The most open ground for research is the improvement of the algorithms themselves. The parameters tuning is a big issue in the use of these algorithms and efforts are being made to reduce the number of parameters that determine the effectiveness of the algorithm. Apart from that, the hybrid algorithms leave a lot of room for research in how the hybridization should be carried out. In some cases, the gbest particle obtained from PSO is used; where as the whole population is mutated in other cases. The mutation operation is sometime applied to a random member of the pbest population where as sometimes on the gbest particle itself.

These different choices affect the effectiveness of the algorithms differently and no fixed convention has been defined. It is up to the researcher to decide and apply his intuition and experience based on trail and error over a number of trials. Thus exploration of these areas in improving the effectiveness of the algorithms based on the best parameters and best approach to hybridization remains to be a work for future research.

In this thesis, a quantum behaved particle swarm optimization was introduced those concepts are radical to the classical concept of swarm optimization. However, it as shown that these algorithms are more effective than the classical PSO. So, it is also a round for future research how new algorithms can be developed by borrowing concepts from different fields of science and applied to improve the existing algorithms. Apart from that, the application of these and other various algorithms in other different kinds of real world applications also remains to be the work for future research This research mainly focused on carrying out simulations on the computer using. So, its implementation on a dedicated digital signal processor (DSP) on real ata can also be looked at in the future. By implementing the digital filters on a DSP with actual data from various sources such as power systems, the ability of the algorithms to actually identify the filter coefficients and design adaptive filters could be tested. On a hardware environment, various other constraints such as memory, storage size, speed of the processor etc. will also come into the effect and hence design of algorithms according to these requirements will pose more challenge to the research.

## **SUMMARY**

In this chapter, summary of all the chapters was covered. The chapter covered the ain motivation of the thesis and briefly summarized how different algorithms are used in two different kinds of optimization problems in the research work. The chapter also concluded that the hybrid algorithms have given better results and also explained the remaining work that can be taken forward for the future research.

## **BIBLIOGRAPHY**

- [1.] Ababneh, J. I., Bataineh, M. H., Linear phase FIR filter design using particle swarm optimization and genetic algorithms, *Digital Signal Processing*, vol. 18, July 2008, pp. 657-668.
- [2.] Chen, S., Mei, T., Luo, M., Yang, X., Identification of Nonlinear System Based on a New Hybrid Gradient-Based PSO Algorithm, *International Conference on Information Acquisition*, July 2007, pp. 265-268.
- [3.] Cai, X., Zhang, N., Venayagamoorthy, G. K., Wunsch, D. C., Time series prediction with recurrent neural networks trained by a hybrid PSO-EA algorithm, *Neurocomputing*, Vol. 70, Issues 13-17, Aug. 2007, pp. 2342-2353
- [4.] el Valle Y., Venayagamoorthy, G. K., Mohagheghi, S., Hernandez, J. C., Harley, R. G., Particle Swarm Optimization: Basic Concepts, Variants and Applications in *Particle Swarm Optimization: Basic Concepts, Variants and Applications* in 2008, pp. 171-195
- [5.] Fang, W., Sun, J., Xu, W., Analysis of Adaptive IIR Filter Design Based on Quantum-behaved Particle Swarm Optimization, *Sixth World Congress on Intelligent Control and Automation*, vol. 1, 2006(a), pp. 3396-3400.
- [6.] Fang, W., Sun, J., Xu, W., Liu, J., FIR Digital Filters Design Based on Quantum-behaved Particle Swarm Optimization, in *Proc. Of the First Intl. Conf. on Innovative Computing, Information and Control*, vol. 1, Aug 2006(b), pp. 615-619.
- [7.] Hao, Z. F., Guo, G. H., Huang, H., A Particle Swarm Optimization Algorithm with Differential Evolution, *International Conference on Machine Learning and Cybernetics*, vol. 2, Aug. 2007, pp. 1031-1035.
- [8.] Hongwei, G., Yanchun L., Identification for Nonlinear Systems Based on Particle Swarm Optimization and Recurrent Neural Network [ultrasonic motor control applications], on *Proc. International Conference on Communications, Circuits and Systems*, vol.2, May 2005.
- [9.] Karaboga, N., Digital IIR Filter Design Using Differential Evolution Algorithm, *EURASIP Journal of Applied Signal Processing* 2005:8, 1269-1276.
- [10.] Karaboga, N., Kalinli, A., Karaboga, D., Designing digital IIR filters using ant colony optimization algorithm, *Engineering Applications of Artificial Intelligence* 17 (2004), 301-309.

# Bi-Level Weighted Histogram Equalization with Adaptive Gamma Correction

Jeena Baby<sup>1</sup>, V. Karunakaran<sup>2</sup>

<sup>1</sup> PG Student, Department of Computer Science and Engineering, Karunya University

<sup>2</sup> Assistant Professor, Department of Computer Science and Engineering, Karunya University

## Abstract:

In this paper the bi-level weighted histogram equalization is combined with adaptive gamma correction method for better brightness preservation and contrast enhancement. The main idea of this method is to initially divide the input dimmed image into R, G and B components and apply the probability density function and weighting constraints on each component separately. And finally, an adaptive gamma correction method is applied to each component and their union produces a brightness preserved and contrast enhanced output image. The performance of this technique is calculated using Absolute mean brightness error (AMBE) measure.

**Keywords:** Contrast enhancement, brightness preservation, histogram equalization, peak signal to noise ratio, absolute mean brightness error, adaptive gamma correction, probability density function, cumulative density function.

## I. INTRODUCTION

Contrast enhancement means improving the visual appearance of the images as well as videos to make it more satisfactory to the human or machine. Contrast enhancement comes under the image enhancement techniques. It is used in both image as well as video processing for better visual perception. Several contrast enhancement techniques are already available. Each technique has got merits and demerits. Histogram equalization is a very traditional technique where the intensity values of the image are redistributed. Due to environmental lighting conditions or because of the defects in the photographic devices, images may suffer from poor contrast. So in order to improve the image quality contrast enhancement is done. Histogram equalization is a simple and effective technique commonly for contrast enhancement [9].

Generally, the image enhancement techniques are categorized into two: direct [2] and indirect enhancement techniques. In direct enhancement techniques, the contrast of the image is directly defined by a definite contrast term [2]. But in indirect enhancement techniques the contrast is improved by redistributing the intensity values of the image [1]. Histogram equalization [9] techniques can be divided as local and global. In global, the active range of intensity can be extended using the histogram of the image and thereby increase the quality. In histogram equalization [9], cumulative distribution function is used to normalize the distribution of intensities, so that the output image will have uniform distribution of intensities. HE will produce a washed out effect in the images [9].

In local HE, the histogram as well as the information obtained from the neighborhood pixels are used for this technique. Here the image is divided into several sub-blocks and then perform HE on each block. The final image is produced by merging these sub-blocks. The most popular indirect enhancement technique is called histogram modification techniques [3]. These are easy techniques which can be implemented in a faster way [1]. A gamma correction method comes under these HM techniques. Here a varying adaptive parameter  $\gamma$  is used. Transform- based gamma correction [17] is the simplest and it can be derived as

$$T(l) = l_{max}(l/l_{max})^\gamma \quad (1)$$

Where  $l$  is the intensity of each pixel in the input image and  $l_{max}$  is the maximum intensity. Since a fixed parameter is used in gamma correction different images will display same changes in intensity. In order to solve this problem a bi level weighted histogram equalization technique proposed in [16] is used. But this method also has some problems like over enhancement.

In section 2, related works are described. Section 3 presents the proposed technique. Section 4 discusses the performance metrics to measure the quality of contrast enhanced image. In section 5, results are discussed and conclusion is given in section 6.

## II. RELATED WORKS

Here some previous works related to histogram equalization and adaptive gamma corrections are discussed. The bi level weighted histogram equalization (BWHE) method [16] segments the input histogram into two sub histograms based on its mean intensity value. The major problem of this method is the over enhancement and the introduction of irregularity, called blocking effect.

In segment dependent dynamic multi-histogram equalization [18] the input histogram is divided into  $n$  segments based either on its mean or median and a range is calculated. Histogram equalization is done based on this range and finally the output image will be normalized. This method is not suitable for color images. Another method called adaptive gamma correction with weighting distribution combines the traditional HE method and TGC method [17].

## III. PROPOSED METHOD

To solve the problems of the earlier works, a new method has been proposed which combines the bi level HE with the adaptive gamma correction method [5]. It will produce a high quality image and the computation is also less. This bi level HE is a technique which combines two methods Weighted Threshold Histogram Equalization (WTHE) [13] and Brightness preserving Bi-Histogram equalization (BBHE) [3]. The algorithm for the proposed method can be described as follows in which the equations are derived from [16] and [17]:

1. Input image is separated into R, G and B components. The following steps are applied to each of the components separately.
2. Compute the probability density function (PDF) of each component.
3. Find the mean pdf  $m$  of each component.
4. Then apply the constraints described below on each component.

$$P_c(r_k) = T(P(r_k)) = \begin{cases} \alpha & \text{if } P(r_k) > \alpha \\ \left(\frac{P(r_k) - \beta}{\alpha - \beta}\right)^r & \text{if } \beta \leq P(r_k) \leq \alpha \\ 0 & \text{if } P(r_k) < \beta \end{cases} \quad (2)$$

Where  $\alpha = v * \max(P(r_k))$ ,  $0.1 < v < 1.0$ ,  $\beta = 0.0001$  and  $r$  is the power factor such that  $0.1 < r < 1.0$ .

5. Find the mean of constrained pdf and then compute the mean error.

$$m_e = m_c - m \quad (3)$$

6. Add the mean error  $m_e$  to constrained pdf  $P_c$ .
7. Find the cumulative density function  $C$  using the  $P_c$ .
8. After that find the weighted pdf value using the below equation:

$$P_w(l) = P_{c_{max}} \left( \frac{P_c(l) - P_{c_{min}}}{P_{c_{max}} - P_{c_{min}}} \right)^\alpha \quad (4)$$

9. Then the modified cdf can be approximated as:

$$C_w(l) = \sum_{i=0}^{l_{max}} P_w(i) / \sum P_w \quad (5)$$



Where the sum of  $pdf_w$  can be calculated as:

$$\sum P_w = \sum_{l=0}^{l_{max}} P_w(l) \quad (6)$$

11. Finally the gamma value is calculated as:

$$\gamma = 1 - C_w(l) \quad (7)$$

And apply this in the gamma correction formula as:

$$T(l) = l_{max} \left( \frac{l}{l_{max}} \right)^\gamma = l_{max} \left( \frac{l}{l_{max}} \right)^{1-C_w(l)} \quad (8)$$

The AGC method will gradually increase the low intensities and avoid the decrease of the high intensity. So in order to avoid the undesirable effects produced in the image, a weighting distribution is also applied [4]. According to [6] and [7] color images are enhanced using HSV color model, where the hue (H) and saturation (S) is used to represent the color content and value (V) represents the luminance intensity.

In bi-level histogram equalization the computation of constrained pdf helps to control the equalization of images. After finding out the pdf it is clamped to an upper threshold value  $\alpha$  and lower threshold value  $\beta$ . The value of  $\nu$  comes in the range of 0.1 to 1.0 so that, the pdf's are clipped with high probabilities. If the value of  $\nu$  is beyond this limit, then over-enhancement occurs. The value of  $r$  is always less so that, over-enhancement is very rare. The mean error is calculated in order to recompense the change in the mean luminance level.

#### IV. IMAGE QUALITY MEASUREMENT

Here two parameters are used to measure the quality of the image. They are: Peak Signal to Noise Ratio (PSNR) for contrast enhancement measurement [8] and Absolute Mean Brightness Error (AMBE) for measuring the mean brightness value [8].

##### 4.1. Peak Signal to Noise Ratio

The PSNR [8] is used to compute the peak signal to noise ratio between two images. The ratio is used as a quality measurement between the original and contrast enhanced image. The higher the PSNR value, the better the quality of the image.

To compute the PSNR value initially the mean-squared error is calculated using the following equation:

$$MSE = \sum_{M,N} [I_1(m,n) - I_2(m,n)]^2 \quad (9)$$

Where M and N are the number of rows and columns in the input images. Then the PSNR can be calculated as follows:

$$PSNR = 10 \log_{10} \left[ \frac{R^2}{MSE} \right] \quad (10)$$

R is the maximum variation in the input image data type.

##### 4.2. Absolute Mean Brightness Error

The proposed method is trying to preserve the brightness of the images by considering the value of AMBE [8]. It is calculated as:

$$AMBE = |E[Y] - E[X]| \quad (11)$$

Where E[Y] is the mean of contrast enhanced image and E[X] is the mean of original image.

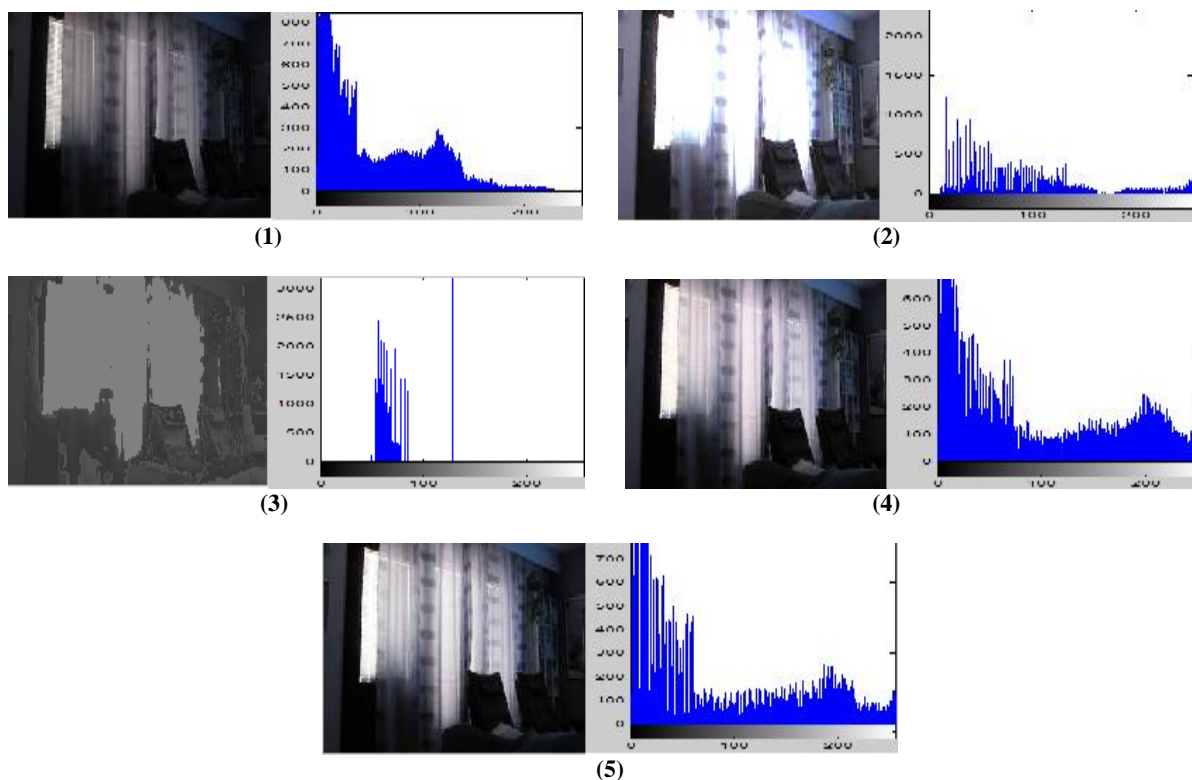
**V. RESULTS AND DISCUSSION**

The performance of the proposed method bi-level histogram equalization with adaptive gamma correction (BWHEAGC) was tested on several color images. The images are shown in Figure 1. To compare the performance of the proposed method the same images are enhanced using AGCWD [17] method, BWHE and SDMHE methods. The performance of all these methods are qualitatively measured using PSNR and AMBE.

In this paper 8 dimmed images are used for contrast enhancement and comparison. The contrast enhancement of images (a) and (e) and their corresponding histogram are shown in Figure 2 and Figure 3 respectively.



**Figure 1. Eight dimmed images**



**Figure 2. Contrast enhancement of image (a), (1) original image, (2) BWHE method, (3) SDMHE method, (4) AGCWD method, (5) Proposed method**

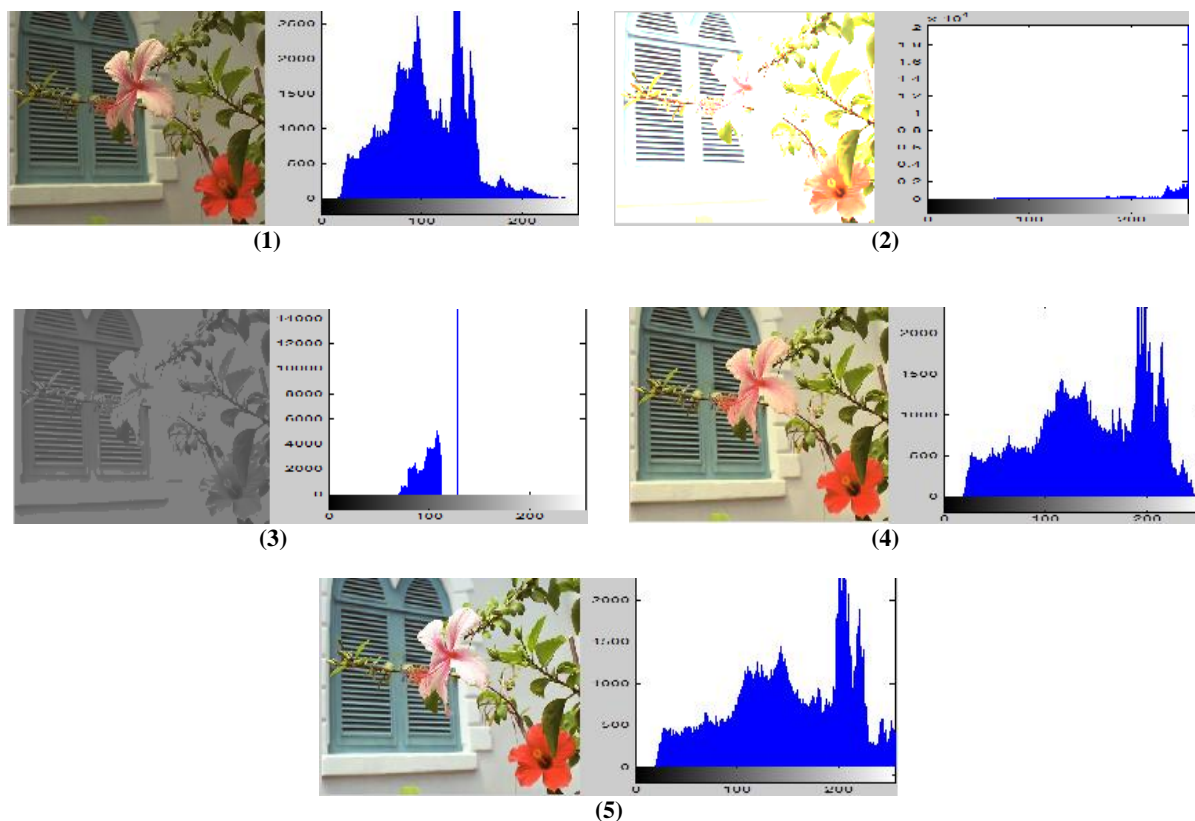


Figure 2. Contrast enhancement of image (e), (1) original image, (2) BWHE method, (3) SDMHE method, (4) AGCWD method, (5) Proposed method

From the above contrast enhanced images it is clear that the BWHE [16] method and the SDMHE [18] method are producing poor results. In BWHE method over-enhancement is the problem. This problem is solved in the proposed method by using an adaptive gamma correction method. The SDMHE method is suitable only for gray level images. The AGCWD [17] method and the proposed method are giving visually acceptable images. But the output produced by the proposed method is better. The performances of these methods are measured using two parameters namely, PSNR and AMBE. The PSNR values are given in Table 1 and the AMBE values are given in Table 2.

Table 1. Comparison of PSNR values

Image \ Method	BWHE	SDMHE	AGCWD	BWHEAGC
(a)	8.2298	10.2595	14.6281	15.9629
(b)	8.3320	11.1200	13.5193	15.2113
(c)	11.9212	7.8042	20.1488	20.2374
(d)	7.8913	10.0006	15.4067	15.6079
(e)	5.4459	14.9751	15.4850	13.4838
(f)	11.0392	11.1143	15.5627	16.7379
(g)	7.9589	9.1970	17.6459	17.6741
(h)	15.1585	9.4252	17.0107	20.2862

By comparing the PSNR values produced by each of the four methods it is clear that the proposed method (BWHEAGC) produces better results. Because, from the definition of PSNR it is clear that, higher its value better will be the image quality. Also from the Table 2, we get the values of AMBE measure which indicates that the brightness is preserved in the output images. Lower the value of AMBE better will be the brightness preservation.

**Table 2. Comparison of AMBE value**

Image \ Method	BWHE	SDMHE	AGCWD	BWHEAGC
(a)	5.8938	0.9116	1.5164	0.4639
(b)	2.3363	9.3737	9.8524	0.1690
(c)	4.2299	0.4463	0.8360	0.0882
(d)	9.6491	3.2908	1.8068	0.5882
(e)	1.0146	0.2724	1.3507	0.6385
(f)	7.0487	6.5929	8.4963	0.3330
(g)	0.1944	8.6451	4.6604	0.4523
(h)	16.5474	3.1738	9.2396	0.4170

## VI. CONCLUSION

In this paper, the contrast of dimmed images are enhanced with the help of bi-level weighted HE with adaptive gamma correction method. This technique is accomplished using two methods, bi-level weighted histogram equalization where, the pdf and cdf is calculated in a constrained manner and the adaptive gamma correction method where, a weighting is done on this constrained pdf. Then using the calculated gamma value the transformation is done. It is computationally simple method and has a high degree of detail preservation. From the calculated values of PSNR and AMBE measures it is clear that the proposed method has better brightness preservation and is the best method for contrast enhancement.

## REFERENCES

- [1] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1921–1935, Sep. 2009.
- [2] A. Beghdadi and A. L. Negrate, "Contrast enhancement technique based on local detection of edges," *Comput. Vis, Graph., Image Process.*, vol. 46, no. 2, pp. 162–174, May 1989.
- [3] Y. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, Feb. 1997.
- [4] M. Kim and M. G. Chung, "Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 54, no. 3, pp. 1389–1397, Aug. 2008.
- [5] Z.-G. Wang, Z.-H. Liang, and C.-L. Liu, "A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP," *Displays*, vol. 30, no. 3, pp. 133–139, Jul. 2009.
- [6] M. Hanmandlu and D. Jha, "An optimal fuzzy system for color image enhancement," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2956–2966, Oct. 2006.
- [7] M. Hanmandlu, O. P. Verma, N. K. Kumar, and M. Kulkarni, "A novel optimal fuzzy system for color image enhancement using bacterial foraging," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 8, pp. 2867–2879, Aug. 2009.
- [8] Kim M. and Chung G., "Recursively Separated and Weighted Histogram Equalization for Brightness Preservation and Contrast Enhancement," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1389-1397, 2008.
- [9] Rafael G. and Richard W., *Digital Image Processing*, Prentice Hall, Gonzalez, 2002.
- [10] Shanmugavadivu P. and Balasubramanian K., "Image Edge and Contrast Enhancement using Unsharp Masking and Constrained Histogram Equalization," *Communications in Computer and Information Science*, vol. 140, no. 2, pp. 129-136, 2011.
- [11] Shanmugavadivu P., Balasubramanian K., and Somasundaram K., "Median Adjusted Constrained PDF based Histogram Equalization for Image Contrast Enhancement," *Communications in Computer and Information Science*, vol. 204, no. 1, pp. 244-253, 2011.
- [12] Shanmugavadivu P., Balasubramanian K., and Somasundaram K., "Modified Histogram Equalization for Image Contrast Enhancement using Particle Swarm Optimization," *International Journal of Computer Science, Engineering and Information Technology*, vol. 1, no. 5, pp. 13-27, 2012.
- [13] Wang Q. and Ward R., "Fast Image/Video Contrast Enhancement Based on Weighted Thresholded Histogram Equalization," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 757-764, 2007.
- [14] Kabir H., Al-Wadud A., and Chae O., "Brightness Preserving Image Contrast Enhancement using Weighted Mixture of Global and Local Transformation Functions," *International Arab Journal of Information Technology*, vol. 7, no. 4, pp. 403-410, 2010.
- [15] A. Polesel, G. Ramponi, and V. Mathews, "Image enhancement via adaptive unsharp masking," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 505–510, Mar. 2000.
- [16] Shanmugavadivu Pichai, Balasubramanian Krishnasamy, and Somasundaram Karuppanagounder, "Bi-Level Weighted Histogram Equalization for Scalable Brightness Preservation and Contrast Enhancement for Images", *The International Arab Journal of Information Technology*, Vol. 10, No. 6, November 2013
- [17] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu, "Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution", *IEEE Trans. Image processing*, Vol. 22, no. 3, Mar 2013.
- [18] MohammadFarhanKhan, EkramKhan,Z.A.Abbasi, "Segment Dependent dynamic multi-histogram Equalization for Image Contrast Enhancement", *Elsevier Digital Signal Processing* 25(2014) 198–223.

# Ultraspherical Solutions for Neutral Functional Differential Equations

A. B. Shamardan<sup>1</sup>, M. H. Farag<sup>1,2</sup>, H. H. Saleh<sup>1</sup>

<sup>1</sup> Mathematics Department, Faculty of Science, Minia University, Egypt

<sup>2</sup> Mathematics and Statistics Department, Faculty of Science, Taif University, Taif, KSA

## ABSTRACT

This paper is concerned with the numerical solution of neutral functional differential equations (NFDEs). Based on the ultraspherical  $\nu$ -stage continuous implicit Runge-Kutta method is proposed. The description and outlines algorithm of the method are introduced. Numerical results are included to confirm the efficiency and accuracy of the method.

**Keywords:** Functional differential equations, Equations of neutral type, Implicit delay equations.

## I. INTRODUCTION

We are interested in the numerical solution of initial-value problem for neutral functional differential equations (NFDEs), which take the form:

$$\begin{aligned} y'(t) &= f(t, y(\cdot), y'(\cdot)) & t_0 < t \leq t_N, \\ y(t) &= g(t) & \alpha \leq t \leq t_0 \end{aligned} \quad (1)$$

where  $g(t) \in C^1[\alpha, t_0]$  and the function  $f(t, y(\cdot), y'(\cdot))$  satisfies the following conditions:-

$H_1$ : For any  $y \in C^1[t_0, t_N]$  the mapping  $t \rightarrow f(t, y(\cdot), y'(\cdot))$  is continuous on  $[t_0, t_N]$ .

$H_2$ : There exist constants  $L_1 \geq 0$ ,  $0 \leq L_2 \leq 1$  such that

$$\begin{aligned} &\left\| f(t, y_1(\cdot), z_1(\cdot)) - f(t, y_2(\cdot), z_2(\cdot)) \right\| \leq \Lambda \\ \Lambda &= L_1 \left\| y_1 - y_2 \right\|_{C^1[\alpha, t_N]} + L_2 \left\| z_1 - z_2 \right\|_{C^0[\alpha, t_N]} \end{aligned}$$

for any  $t \in [t_0, t_N]$ ,  $y_1, y_2 \in C^1[\alpha, t_N]$  and  $z_1, z_2 \in C^0[\alpha, t_N]$ .

Under the conditions  $H_1$  and  $H_2$  the problem (1) has a unique solution  $y(x)$  [1]. The equations of type (1) have applications in many fields such as control theory, oscillation theory, electrodynamics, biomathematics, and medical science. Numerical methods for the problem (1) were discussed extensively by many authors; see [2-26].

This paper is concerned with the numerical solution of neutral functional differential equations (NFDEs). Based on the ultraspherical  $\nu$ -stage continuous implicit Runge-Kutta method is proposed. In section 2, we will adapt a finite

Ultraspherical expansion to approximate  $\int_{t_i}^t f(s) ds$  and  $f(\cdot)$  on the interval  $I_i, i = 0(1)N - 1$ . Also, an easily implemented numerical method for NFDEs will be derived. Finally, in section 3 we present some numerical examples; which show that the presented method provides a noticeable improvement in the efficiency over some previously suggested methods.

## II. THE NUMERICAL METHOD

### 2.1 The Description of the method

Let  $\Delta := (t_0 < t_1 < \dots < t_N)$  define a partition for  $[t_0, t_N]$ , with the step size  $h_i = t_{i+1} - t_i$ .

Each subinterval  $I_i = [t_i, t_{i+1}]$  is divided by the Chebyshev collocation points:

$$t_{ij} = t_i + h_i \xi_j, \xi_j = \frac{1}{2} (1 - \cos \frac{j\pi}{\nu}), j = 0(1)\nu \tag{2}$$

This method is based on a finite Ultraspherical expansion in each subinterval  $I_i, i = 0(1)N - 1$ . Consider the approximation  $\tilde{f}(t)$  of  $f(t)$  for  $t = t_i + \theta h_i, 0 \leq \theta \leq 1, i = 0(1)N - 1$  as follows:

$$\tilde{f}(t) = \sum_{r=0}^{\nu} a_r C_r^{[\alpha]} \left( \frac{2(t-t_i)}{h_i} - 1 \right), t \in [t_i, t_{i+1}] \tag{3}$$

where

$$a_r = \frac{\pi}{\nu \lambda_r^{[\alpha]}} \sum_{l=0}^{\nu} \tilde{f}(t_{il}) \left( \sin \left( \frac{l\pi}{\nu} \right) \right)^{2\alpha} C_r^{[\alpha]} \left( -\cos \left( \frac{l\pi}{\nu} \right) \right) \tag{4}$$

$$t_{il} = t_i + \frac{h_i}{2} (1 - \cos \left( \frac{l\pi}{\nu} \right))$$

Here  $C_r^{[\alpha]}(x)$  is the r-th Ultraspherical polynomial. As especial cases, at  $\alpha = 0$  give Chebyshev Polynomials of the first kind  $C_r^{[0]}(x) = T_r(x)$ , at  $\alpha = \frac{1}{2}$  give Legendre Polynomials

$C_r^{[\frac{1}{2}]}(x) = P_r(x)$ , at  $\alpha = 1$  give Chebyshev Polynomials of the second kind  $C_r^{[1]}(x) = \frac{1}{1+r} R_r(x)$ .

A summation symbol with double prims denotes a sum with the first and last terms halved.

Now, we can easily show that the following relations are true:

$$I_r(\theta) = \int_{-1}^{2\theta-1} C_r^{[\alpha]}(s) ds = \begin{cases} C_1^{[\alpha]}(2\theta-1) + 1 & \text{if } r = 0 \\ \frac{(2\alpha+1)}{4(\alpha+1)} [C_2^{[\alpha]}(2\theta-1) - 1] & \text{if } r = 1 \\ \frac{1}{2(\alpha+r)} \left( \frac{2\alpha+r}{r+1} \Phi_1 - \frac{r}{2\alpha+r-1} \Phi_2 \right) & \text{if } r \geq 2 \end{cases} \quad (5)$$

where  $\Phi_1 = C_{r+1}^{[\alpha]}(2\theta-1) + (-1)^r$ ,  $\Phi_2 = C_{r-1}^{[\alpha]}(2\theta-1) + (-1)^r$

Using (3), (4) and (5) the indefinite integral  $\int_{t_i}^{t_i + \theta h_i} f(s) ds$  takes the form

$$\int_{t_i}^{t_i + \theta h_i} f(s) ds = \frac{h_i}{2} \sum_{r=0}^{\nu} b_l^{[\alpha]}(\theta) f(t_{il}) \quad (6)$$

where

$$b_l^{[\alpha]}(\theta) = \frac{\pi \left[ \sin \left( \frac{l\pi}{\nu} \right) \right]^{2\alpha}}{2(1 + \delta_{l0} + \delta_{l\nu})} \sum_{l=0}^{\nu} \frac{I_r(\theta) C_r^{[\alpha]} \left( -\cos \left( \frac{l\pi}{\nu} \right) \right)}{\lambda_r^{[\alpha]}}, \quad l = 0(1)\nu \quad (7)$$

and  $\delta_{l\nu}$  is the Kronecker delta.

From the relations (2), (6) and (7), we obtained the following results:

$$\left[ \int_{t_i}^{t_{ij}} f(s) ds \right] = \frac{h_i}{2} B^{[\alpha]} F_i \quad (8)$$

$$\left[ \int_{\underbrace{t_i \dots t_i}_{k \text{ - times}}}^{t_{ij}} \dots \int_{t_i}^{t_{ij}} \tilde{f}(s) ds \right] = \left( \frac{h_i}{2} \right)^k (B^{[\alpha]})^k F_i = \frac{\left( \frac{h_i}{2} \right)^k}{(k-1)!} B^{[\alpha]} K F_i \quad (9)$$

$$\int_{t_i}^{t_{ij}} \tilde{f}(s, t) ds = \sum_{l=0}^{i-1} \frac{h_l}{2} \sum_{s=0}^{\nu} b_{vs}^{[\alpha]} \tilde{f}(t, t_{ls}) + \frac{h_i}{2} \sum_{s=0}^{\nu} b_{js}^{[\alpha]} \tilde{f}(t, t_{is}), \quad j = 0(1)\nu \quad (10)$$

where

$$B^{[\alpha]} = [b_{js}^{[\alpha]}]_{j,s=0}^{\nu}, b_{js}^{[\alpha]} = b_s^{[\alpha]}(c_j), F_i = [\tilde{f}(t_{i0}), \dots, \tilde{f}(t_{i\nu})]^T$$

and  $K = \text{diag} [(c_j - c_0)^{k-1}, \dots, (c_j - c_{\nu})^{k-1}]$ .

We notice that the matrix  $B$  of El-Gendi's method [27], is the same matrix  $B^{[\alpha]} = [2 b_{js}^{[0]}]_{j,s=0}^{\nu}$ . On the interval  $(t_i, t_{i+1}]$ , rewriting (1) in the form:

$$\tilde{y}(t) = \tilde{y}(t_i) \int_{t_i}^t \tilde{z}(s) ds, \quad \tilde{z}(t) = f(t, \tilde{y}(\cdot), \tilde{z}(\cdot)) \quad (11)$$

Suppose the approximations of  $\tilde{y}(t)$  and  $\tilde{z}(t)$  are given for  $t \leq t_i$ . On  $(t_i, t_{i+1}]$ , we define the  $\nu$ -stage method  $[\nu - UM]$ , as follows

$$\tilde{z}(t_{il}) = f(t_{il}, \tilde{y}(\cdot), \tilde{z}(\cdot)), \quad l = 1(1)\nu \quad (12)$$

where

$$\tilde{y}(t) = \begin{cases} g(t) & \text{if } t \leq t_0 \\ \tilde{y}(t_{j-1}, \nu) + \frac{h_j}{2} b_0^{[\alpha]}(\theta) \tilde{z}(t_{j-1}, \nu) + \frac{h_j}{2} \sum_{l=1}^{\nu} b_l^{[\alpha]}(\theta) \tilde{z}(t_{jl}) & \text{if } t = t_j + \theta h_j \in I_j \end{cases} \quad (13)$$



$$\tilde{z}(t) = \begin{cases} g'(t) & \text{if } t \leq t_0 \\ \sum_{r=0}^{\nu} a_r C_r^{[\alpha]} \left( \frac{2(t-t_j)}{h_j} - 1 \right) & \text{if } t = t_j + \theta h_j \in I_j \end{cases} \quad (14)$$

$$a_r = \frac{\pi}{\nu \lambda_r^{[\alpha]}} \sum_{l=0}^{\nu} \tilde{z}(t_{jl}) \left( \sin \left( \frac{l\pi}{\nu} \right) \right)^{2\alpha} C_r^{[\alpha]} \left( -\cos \left( \frac{l\pi}{\nu} \right) \right) \quad (15)$$

### 2.2 The Algorithm of $\nu$ -stage method

The described method represents a generalization of the methods given by Jackiewicz [7-8], for  $\nu = 1, 2, 3$  and  $\alpha = 0$ . The algorithm of the  $\nu$ -stage method is given below:-

**STEP 1:** Input  $N, h_j, I_j, j = 0(1)N - 1$ .

**STEP 2:** Put  $i = 0$

**STEP 3:** Compute  $\tilde{y}(t_{ij})$  and  $\tilde{z}(t_{ij})$  on  $I_j, j = 1(1)N - 1$ , by solving the system of  $\nu$ -equation (12).

**STEP 4:** Store the computed values  $\tilde{y}(t_{ij})$  and  $\tilde{z}(t_{ij})$  on  $I_j, j = 1(1)N - 1$ .

**STEP 5:** If  $i = N - 1$  go to step 6, otherwise set  $i = i + 1$  and go to step 3.

**STEP 6:** Output the results  $\tilde{y}(t_{ij})$  and  $\tilde{z}(t_{ij}), i = 0(1)N - 1$ .

### III. NUMERICAL EXAMPLES

In this section, we present the result of some computational experiments by applying our  $\nu - UM$  method.

#### Example 1: (Jackiewicz [10])

$$y'(t) = e^{-y(t)} + \sin(y'(\beta(t))) - \sin\left(\frac{1}{3 + \beta(t)}\right), \quad t \in (0, 10],$$

$$y(0) = \ln(3), \quad y'(0) = \frac{1}{3} \quad (16)$$

Here  $\beta(t) = 0.5 t (1 - \cos(2\pi t))$ . The exact solution is  $y(t) = \ln(3 + t)$ .

**Example 2: (Jackiewicz [10])**

$$y'(t) = \frac{\sin(t y'(t)) - \sin(e^{y(t)})}{16} + \frac{1}{t}, \quad t \in (1,4],$$

$$y(1) = 0, \quad y'(1) = 1$$
(17)

Here the exact solution is  $y(t) = \ln(t)$ .

Table 1			Table 2	
	<b>Example 1: <math>y(10) = 2.5649494</math></b>		<b>Example 2: <math>y(4) = 1.386294361</math></b>	
$h$	$2-UM$	$C_0C_1$	$2-UM$	$C_0C_1$
$2^{-1}$	1.21E-4	-2.5E-4	-1.12E-4	-4.00E-1
$2^{-2}$	7.62E-6	-6.25E-2	-8.00E-6	-7.00E-2
$2^{-3}$	4.76E-7	6.25E-2	5.00E-7	-5.00E-3
$2^{-4}$	2.98E-8	7.42E-2	-3.13E-8	3.13E-3
$2^{-5}$	1.86E-9	2.93E-3	-1.95E-9	-3.13E-4
$2^{-6}$	1.16E-10	-7.32E-4	-1.22E-10	-3.13E-4
$2^{-7}$	1.02E-11	-7.93E-4	-7.63E-12	-8.79E-5
$2^{-8}$	-7.10E-12	-3.66E-4	-4.77E-13	-2.44E-5
$2^{-9}$	1.15E-11	-8.39E-5	-2.98E-14	-6.10E-6
$2^{-10}$	-4.88E-12	-5.72E-6	-4.66E-15	-1.53E-6

**Example 3: (Jackiewicz [10])**

$$y'(t) = e^{1-2t^2} y(t^2) \left[ y' \left( \frac{t-1}{t+1} \right) \right]^{(1+t)}, \quad t \in (0,1],$$

$$y(t) = e^t, \quad t \in [-1, 0]$$
(18)

Here the exact solution is  $y(t) = e^t$ .

**Example 4: (Jackiewicz [10])**

$$y'(t) = 2 \cos(2t) \left( y \left( \frac{t}{2} \right) \right)^{2 \cos(t)} + \ln \left( \frac{y' \left( \frac{t}{2} \right)}{2 \cos(t)} \right) - \sin(t), \quad t \in (0,1],$$

$$y(0) = 1, \quad y'(0) = 2$$
(19)

Here the exact solution is  $y(t) = e^{\sin(2t)}$ .

Table 3			Table 4	
	Example 3: $y(1) = 2.7182818$		Example 4: $y(1) = 2.4825777$	
$h$	2-UM	$C_0C_1$	2-UM	$C_0C_1$
$2^{-1}$	1.25E-3	2.75E-2	-5.94E-2	4.28E-1
$2^{-2}$	7.81E-5	-2.50E-3	-4.41E-3	5.44E-2
$2^{-3}$	4.88E-6	-2.03E-3	-3.08E-4	1.91E-2
$2^{-4}$	3.05E-7	-6.64E-4	-2.11E-5	5.59E-3
$2^{-5}$	2.86E-8	-1.86E-4	-1.42E-6	1.33E-3
$2^{-6}$	1.79E-9	-4.88E-5	-9.30E-8	3.13E-4
$2^{-7}$	7.45E-11	-1.22E-5	-5.96E-9	7.45E-5
$2^{-8}$	9.31E-12	-3.20E-6	-3.73E-10	1.82E-5
$2^{-9}$	7.28E-13	-8.01E-7	-2.33E-11	4.46E-6
$2^{-10}$	-2.73E-15	-2.00E-7	-1.40E-12	1.12E-6

In tables (1)-(4), we give  $\mathbf{E}$  for 2-UM method and  $\mathbf{E}$  for the best method of Jackiewicz;  $C_0 C_1$  method [10], where  $\mathbf{E}$  denotes the global error at the end point  $t_N$ .

**Example 5: (Kappel-Kunish [15] and Jackiewicz [10])**

$$\begin{aligned}
 y'(t) &= y(t) + y(t-1) - 0.25 y'(t-1), \quad t \in (0, 2], \\
 y(t) &= -t, \quad y'(t) = -1, \quad t \in [-1, 0]
 \end{aligned}
 \tag{20}$$

Here the exact solution is given by

$$\begin{aligned}
 y(t) &= t + 0.25 e^t - 0.25, \quad t \in [0, 1] \\
 y(t) &= -t + 0.25 e^t + 0.5 + \frac{17 + 3t}{16} e^{(t-1)}, \quad t \in [1, 2]
 \end{aligned}$$

Here, the first derivative has a discontinuity at  $t = 0$  and  $t = 1$ .

In example 5, we give  $\mathbf{E}$  for the method 2-UM, 5-UM, 8-UM and  $C_0 C_1$  in Table (5), the  $\mathcal{V}$ -UM methods, for large  $\mathcal{V}$ , make a little improvement in the computed results, the reason is due to there exist discontinuity for the first derivative of  $y(t)$  at  $t = 0$  and  $t = 1$ . These results indicate that the  $\mathcal{V}$ -UM method is better than the one-step methods of Jackiewicz [10].

Table 5				
Example 5: $y(2) = 4.2547942$				
$h$	2 – UM	5 – UM	8 – UM	$C_0C_1$
$2^{-1}$	1.6	2.03E-1	8.00E-2	6.53
$2^{-2}$	8.15E-1	1.00E-1	4.00E-2	4.22
$2^{-3}$	4.10E-1	4.95E-2	2.00E-2	2.35
$2^{-4}$	2.05E-1	2.48E-2	1.00E-2	1.22
$2^{-5}$	1.03E-1	1.23E-2	5.00E-3	6.14E-1
$2^{-6}$	5.13E-2	6.16E-3	2.50E-3	3.08E-1
$2^{-7}$	2.56E-2	3.08E-3	1.25E-3	1.54E-1
$2^{-8}$	1.28E-2	1.54E-3	6.25E-4	7.69E-2
$2^{-9}$	6.41E-3	7.70E-4	3.13E-4	3.84E-2
$2^{-10}$	3.20E-3	7.70E-4	1.56E-4	1.92E-2

**Example 6:** (Castleton-Grimm [4] and Jackiewicz [8])

$$y'(t) = \frac{-4 t y^2(t)}{(\ln(\cos(2t)))^2 + 4} + \tan(2t) + \frac{\tan^{-1}(z(t))}{2}, \quad t \in (0, 0.75],$$

$$y(0) = 0, \quad y'(0) = 0$$
(21)

where  $z(t) = y' \left( \frac{t y^2}{1 + y^2(t)} \right)$ . The theoretical solution is  $y(t) = -0.5 \ln(\cos(2t))$ .

**Example 7:** (Castleton-Grimm [4] and Jackiewicz [8])

$$y'(t) = (1 + u(t)) \cos(t) + y(t) z(t) - \sin(t + t \sin^2(t)), \quad t \in (0, 1],$$

$$y(0) = 0, \quad y'(0) = 1$$
(22)

where  $u(t) = y(t y^2(t))$ ,  $z(t) = y'(t y^2(t))$ .

The theoretical solution is  $y(t) = \sin(t)$ .

**Example 8:** (Pouzet [16] and Jackiewicz [8])

$$y'(t) = e^{-t^2} (1 + t^2) + t - y(t) + \int_0^t t^2 e^{-st} y(s) y'(s) ds, \quad y(0) = 0, \quad t \in (0, 2]$$
(23)

The exact solution is  $y(t) = t$ .

Table 6 : Example 6 (Castleton-Grimm [4] and Jackiewicz [8])

$t_n$	$y(t_n)$	$y_n(2^{-4})$	$y_n(2^{-6})$	$y_n(2^{-8})$	$y_n(2^{-10})$
0.0625	0.003916465	0.003916470	0.003926723	0.003916465	0.003917107
		0.003916465	0.003916505	0.003916465	0.003916468
0.1250	0.015790526	0.015790547	0.015832887	0.015790526	0.015793178
		0.015790526	0.015790691	0.015790526	0.015790536
0.1875	0.036012506	0.036012544	0.036113197	0.036012505	0.036018812
		0.036012506	0.036012900	0.036012506	0.036012531
0.2500	0.065292120	0.065292132	0.065486194	0.065292114	0.065304276
		0.065292120	0.065292880	0.065292120	0.065292168
0.3125	0.104766351	0.104766181	0.105105260	0.104766333	0.104787582
		0.104766350	0.104767677	0.104766351	0.104766433
0.3750	0.156199949	0.156199327	0.156765982	0.156199923	0.156235405
		0.156199949	0.156202162	0.156199949	0.156200087
0.4375	0.222365362	0.222364313	0.223301956	0.222365342	0.222423942
		0.222365362	0.222369022	0.222365362	0.222365591
0.5000	0.307813235	0.307813411	0.309395557	0.307813221	0.307912401
		0.307813235	0.307819432	0.307813235	0.307813622
0.5625	0.420618861	0.420623077	0.423464892	0.420618861	0.420797810
		0.420618860	0.420630041	0.420618861	0.420619559
0.6250	0.577079895	0.577100894	0.582906497	0.577079967	0.577448783
		0.577079896	0.577102963	0.577079895	0.577081337
0.6875	0.818538930	0.818682074	0.834410205	0.818539584	0.819566071
		0.818538933	0.818603259	0.818538930	0.818542950
0.7500	1.324391827	1.328895912	1.430890306	1.324428196	1.332263471
		1.324391983	1.324892390	1.324391828	1.324423149

Table 7 : Example 7 (Castleton-Grimm [4] and Jackiewicz [8])

$y(t_n = 0.25) = 0.247403959,$ $y(t_n = 0.50) = 0.479425539$ $y(t_n = 0.75) = 0.681638760,$ $y(t_n = 1.0) = 0.841470985$					
$t_n$	$y_n(2^{-2})$	$y_n(2^{-4})$	$y_n(2^{-6})$	$y_n(2^{-8})$	$y_n(2^{-10})$
0.25	0.247404002	0.245934360	0.247403959	0.247306957	0.247403959
	0.247397577	0.247403959	0.247403556	0.247403959	0.247403934
0.50	0.479423742	0.475355054	0.479425535	0.479158061	0.479425539
	0.479408626	0.479425539	0.479424471	0.479425539	0.479425472
0.75	0.681635238	0.672679530	0.681638752	0.681051318	0.681638760
	0.681602967	0.681638760	0.681636491	0.681638760	0.681638618
1.00	0.841459232	0.825030283	0.841470963	0.840379343	0.841470984
	0.841404641	0.841470985	0.841466775	0.841470985	0.841470722

Table 8 :Example 8 (Pouzet [16] and Jackiewicz [8])

$t_n$	$y(t_n)$	$y_n(0.1)$	$y_n(0.05)$	$y_n(0.025)$	$y_n(0.0125)$
0.1	0.100000000	0.100000000 0.099999999	0.100000000 0.099999999	0.100000000 0.100000000	0.100000000 0.100000000
0.2	0.200000000	0.200000000 0.199998855	0.200000000 0.19999972	0.200000000 0.19999993	0.200000000 0.19999998
0.3	0.300000000	0.300000000 0.299999115	0.300000000 0.29999806	0.300000000 0.29999958	0.300000000 0.29999988
0.4	0.400000000	0.399999999 0.399996708	0.400000000 0.399999241	0.400000000 0.399999815	0.400000000 0.39999954
0.5	0.500000000	0.499999998 0.499990890	0.500000000 0.499997844	0.500000000 0.499999469	0.500000000 0.499999868
0.6	0.600000000	0.599999994 0.599979283	0.600000000 0.599995019	0.600000000 0.599998768	0.600000000 0.599999693
0.7	0.700000000	0.699999989 0.699959044	0.699999999 0.699990053	0.700000000 0.699997533	0.700000000 0.699999384
0.8	0.800000000	0.799999981 0.799927095	0.799999999 0.799982170	0.800000000 0.799995569	0.800000000 0.799998894
0.9	0.900000000	0.899999953 0.899880395	0.899999998 0.899970604	0.900000000 0.899992684	0.900000000 0.899998173
1.0	1.000000000	0.999999953 0.999816214	0.999999997 0.999954664	1.000000000 0.999988706	1.000000000 0.999997179
1.1	1.100000000	1.099999933 1.099732385	1.099999996 1.099933803	1.100000000 1.099983497	1.100000000 1.099995877
1.2	1.200000000	1.199999907 1.199627509	1.199999994 1.199907665	1.200000000 1.199976968	1.200000000 1.199994245
1.3	1.300000000	1.299999876 1.299501081	1.299999992 1.299876121	1.300000000 1.299969086	1.300000000 1.299992275
1.4	1.400000000	1.399999839 1.399353516	1.399999990 1.399839272	1.399999999 1.399959876	1.400000000 1.399989973
1.5	1.500000000	1.499999794 1.499186056	1.499999987 1.499797427	1.499999999 1.499949416	1.500000000 1.499987358
1.6	1.600000000	1.599999740 1.599000583	1.599999984 1.599751058	1.599999999 1.599937824	1.600000000 1.599984460
1.7	1.700000000	1.699999675 1.698799343	1.699999980 1.699700726	1.699999999 1.699925339	1.700000000 1.699981314
1.8	1.800000000	1.799999596 1.798584660	1.799999975 1.799647014	1.799999998 1.799911809	1.800000000 1.799977956
1.9	1.900000000	1.899999502 1.898358666	1.899999969 1.899590455	1.899999998 1.899897665	1.900000000 1.899974420
2.0	2.000000000	1.999999389 1.998123103	1.999999962 1.999531488	1.999999998 1.999882918	2.000000000 1.999970733

In Tables (6)-(8), we give the exact solution  $y(t_n)$  in the second column, and  $y_n(h)$  which denotes the computed value  $y_h(t_n)$  at every  $t = t_n$ , the computed value of  $y_n(h)$  are represented in two lines, the first line for **2-UM** and the second line for the method of Jackiewicz [8]. The results given in Tables (6) and (7) are much better than those obtained by Castleton-Grimm [4] and also, those obtained by Jackiewicz [8]. The results given in Table (8) are much better than those obtained by Jackiewicz [8].

When solving the nonlinear equations, the computations are terminated when two successive approximations differed by less than  $10^{-3} h_3$ .

#### IV. CONCLUSIONS




In this paper we construct a method based on the Ultraspherical approximation. This method can be applied to solve different types of NFDEs. The experimental comparison, presented in this paper, shows that this method is more efficient than the previously introduced methods. In addition, the **V-UM** method can be easily implemented on computer compared with the Lagrange multipliers and their integrals which given by Jackiewicz [10].

#### References

- [1] Z. Kamont, M. Kwapisz, On the Cauchy problem for differential delay equations in a Banach space, *Math. Nachr.*, 74: 173-190, 1976.
- [2] A. Bellen, Constrained mesh methods for functional differential equations, *ISNM*, 74:52-70, 1970.
- [3] H. Brunner, The numerical solution of the neutral Volterra integro-differential equations with delay arguments, *Annals of Numerical Mathematics*, 1: 309-322, 1994.
- [4] R. N. Castleton, A first order method for differential equations of neutral type, *Math. Comput.* 27: 571-577, 1973.
- [5] C. W. Cryer, Numerical methods for functional differential equations In: *Delay and functional differential equations and their applications* (K. Schmitt, ed.), 17-101. New York: Academic Press, 1972.
- [6] U. Hornung, Euler-Verfahren fur neutral functional differential glcichungen, *Numer. Math.*, 24:233-240, 1975.
- [7] Z. Jackiewicz, One-step methods for the numerical solution of Volterra functional differential equations of neutral type, *Applicable anal.*, 12: 1-11, 1981.
- [8] Z. Jackiewicz, The numerical solution of Volterra functional differential equations of neutral type}, *SIAM J. Numer. Anal.* 18: 615-626, 1981.
- [9] Z. Jackiewicz, Adams methods for neutral functional differential equations, *Numer. Math.* 39: 221-230, 1982.
- [10] Z. Jackiewicz, One-step methods of any order for neutral functional differential equations, *SIAM J. Numer. Anal.* 21: 486-511, 1981.
- [11] Z. Jackiewicz, Quasilinear multistep methods and variable-step predictor-corrector methods for neutral functional differential equations, *SIAM J. Numer. Anal.*, 23: 423-452, 1986.
- [12] Z. Jackiewicz, One-step methods for neutral delay differential equations with state dependent delays, *Numerical Analysis Technical Report 651, 05-2*. University of Arkansas, Fayetteville, 1985.
- [13] Z. Jackiewicz, The numerical solution of functional differential equations, a survey, *Mat. Stos.* , 33: 57-78, 1991.
- [14] Z. Jackiewicz, The numerical integration of NFDEs by fully implicit One-step methods}, *ZAMM*, 75:207-221, 1995.
- [15] F. Kappel, K. Kunisch, Spline approximations for neutral functional differential equations, *SIAM J. Numer. Anal.*, 18:1058-1080,1981.
- [16] P. Pouzet, M'ethode d'intgration num'erique des e'quations inte'grales et inte'gro-diffe'rentielles du type de Volterra de seconde espe'ce. Formules de Runge-Kutta, *Symposium on the Numerical Treatment of Ordinary Differential Equations, Integral and Integro-differential Equations*, (Rome, 1960), Birkha"user Verlag, Basel, 362-368,1960.
- [17] M. Zennaro, Natural continuous extensions of Runge-Kutta methods, *Math. Comput.* 46 (1986) 119-133.

- [18] M. Zennaro, Delay differential equations theory and numerics, In: Theory and Numerics of Ordinary and Partial Differential Equations, Clarendon Press, Oxford, 291-333,1995.
- [19] T. S. Zverkina, A modification of finite difference methods for integrating ordinary differential equations with nonsmooth solutions, (in Russian). Z. Vycisl Mat. i Mat. Fiz.[Suppl.] ,4 : 149-160, 1994.
- [20] W. S. Wang, S. Fu Li, K. Su, Nonlinear stability of Runge-Kutta methods for neutral delay differential equations), J. Comput. Appl. Math. , 214: 175-185, 2008.
- [21] Y. Qu, M. Y. Li, J. Wei, Bufurcation analysis in a neutral differential equation, J. Math. Anal. Appl., 378: 387-402, 2011.
- [22] J. J. Zhao, Y. Xu, M. Z. Liu, Stability analysis of numerical methods for linear neutral Volterra delay-integro-differential system, J. Appl. Math. Comput, 167:1062-1079, 2005.
- [23] Z. Jackiewicz, E. Lo, Numerical solution of neutral functional differential equations by Adams methods in divided difference form, J. Comput. Appl. Math. , 189: 592-605, 2006.
- [24] W. S. Wang, S. Fu Li, K. Su, Nonlinear stability of general linear methods for neutral delay differential equations, J. Comput. Appl. Math. , 214: 592-601, 2009.
- [25] A. Bellen, N. Guglielmi, M. Zennaro, Numerical stability of nonlinear delay differential equations of neutral type, J. Comput. Appl. Math., 125 : 251-263,2000.
- [26] M. A. Hussien, Chebyshev solution for neutral delay differential equations, Inter. J. Computer Math., 73: 253-260, 1999.
- [27] S. E. El-Gendi, Chebyshev solution of differential, integral and integro-differential equations, Comput. J., 12: 282-287, 1969.

**Author names and affiliations**

	<p><b>A. B. SHAMARDAN- Professor of Mathematics (Numerical solutions of Delay and Neutral functional differential equations), Mathematics Department, Faculty of Science, Minia University, Minia, Egypt.</b>  <i>E-mail: shmrndn52@miniauniv.edu.eg</i></p>
	<p><b>M. H. FARAG – Professor of Mathematics (Numerical Analysis and Optimal control PDEs), Mathematics and Statistics Department, Faculty of Science, Taif University, Hawia (P.O. 888), Kingdom of Saudi Arabia (KSA). <u>OR</u> Mathematics Department, Faculty of Science, Minia University, Minia, Egypt.</b>  <i>E-mail: farag5358@yahoo.com</i></p>
	<p><b>H. H. SALEH - Student PhD (Numerical Analysis - Delay and Neutral functional differential equations - Optimal control PDEs), Mathematics Department, Faculty of Science, Minia University, Minia, Egypt.</b>  <i>E-mail: heba_saleh16@yahoo.com</i></p>



# Solid Waste Management in Mahaboobnagar Municipality

<sup>1</sup>, Dr. C.Sarala, <sup>2</sup>, G.SreeLakshmi,

<sup>1</sup>, Associate Professor, Centre for Water Resources Institute of Science and Technology, JNTU Hyderabad.

<sup>2</sup>, Research Scholar, Centre for Water Resources Institute of Science and Technology, JNTU Hyderabad.

## ABSTRACT:

The 74<sup>th</sup> Amendment of the constitution of India in 1999, made Municipal authorities in the country as a thier tier to government. The 12<sup>th</sup> schedule of the constitution envisaged functions to be performed by t municipal authorities, one among these functions is solid waste management. The Ministry of Environme and Forest has notified municipal solid waste rules 2000 under the Environment Protection Act 198 According to these rules all municipal authorities were expected to improve solid waste manageme practices in terms of a fore said rules by 2003, but the situation did not improve as expected for want adequate technical knowledge, there is no proper management facility for solid waste. Uncontrolled dumpi of municipal solid waste has been observed at the road side. There is no processing facility or dispos practices in any urban local bodies. The biomedical slaughter house waste is getting mixed with solid was and altering the characteristics of waste hence there is a need to develop a proper management systems. Th paper presents the waste management system of Mahabubnagar municipality in Andra Pradesh State of Ind to implement municipal solid waste rules 2000. Expediitiously in Mahabubnagar municipality by the proce of modernizing the system of solid waste management.

**Key words:** solid waste management,emissions,land filling,biological process, green house gases.

## I. INTRODUCTION:

At 1992 Rio Earth summit, countries agreed to the UN(United Nations) framework convention on climate change (UNFCCC) in response to growing evidence that human activity was contributing to global warming. The UNFCCC contained a non binding commitment by industrialized countries that they would reduce their emission of green houses to 1990 levels by the year 2000. It soon became clear that this wasn't enough to avert the dangerous climate change and in1995, at the first conference of parties (COP1) after the convention came into force, parties began to negotiate a protocol that would set tighter and legally binding targets for reducing green house gas emissions. In 1997 at the third COP the convention at Kyoto in Japan, parties agreed on a protocol that set target for industrial used countries to reduce their emissions by an average of 5.2% below 1990 levels in the period 2008-2012 known as the first commitment period.

The Kyoto protocol of 1997 is crucial step in the implementation of the United Nations framework convention on climate change as it sets legally binding emission targets for a basket of six green house gases. A market mechanism called emission trading was established under the Kyoto protocol which allows governments or private entities in the industrialized countries to implement machine reduction projects and receives and credit in the form of certified emission reductions (CER) also called carbon credits. In the Kyoto protocol the developed countries committed themselves to reduce their GHG emissions by 5.2% by 2012. To reach Kyoto protocol allowed three flexibility mechanisms, they are Joint Implementation (JI), Clean Development Mechanism (CDM), and International Emission Trading (IET). These are defined wide article 6, 12 & 17 of Kyoto protocol respectively. The clean development mechanism is a project based mechanism to assist developing countries in meeting the targets to reduce GHG in achieving their sustainable development objectives such CDM products would also lead to indirect benefits in the host countries like Income generation, Improvement more slightly to prove to be a primary funding source for climate change mitigation projects in developing countries as a part of multi-billion dollar Green House mitigation market.

Clean Development Mechanism Eligible Projects:

- Hydro power plants
- Tree Plantation, also using genetically modified tree
- Wind farms
- Solar Energy Projects
- Geothermal energy projects
- Biomass energy projects
- Waste incineration projects
- Projects Reducing emission of other Green House gases.

Methane and carbon dioxide are the emissions of solid waste dumping sites. Hence the option of composting of municipal solid waste is chosen to adhere to the clean development mechanism of Kyoto protocol.

## II. METHODOLOGIES in MUNICIPAL SOLID WASTE PROCESSING TECHNIQS:

One of the most important aim of municipal waste management is the safe disposal of waste generated daily this would involve separation of recyclable fraction and recycling the same, beneficial utilization of organic fraction of the waste and disposal of inert into the landfill.

There are several municipal solid waste processing technologies which are being followed in various parts of the world. Besides source reduction, reuse and recycling broad categories of available technologies for processing municipal solid waste.

**Table: I**

MUNUCIPAL SOLID WASTE PROCESSING TECHNIQUES	
Waste processing technology group	Waste processing technology
Thermal processing technologies	Incineration
	Pyrolysis
Biological processing techniques	
	Anaerobic digestion(bio-methanation)
Physical processing technologies	Pyrolysis\gasification
	Plasma arc gasification
	Aerobic digestion(composting)
	Size reduction

Final functional element in solid waste management system is treatment and disposal the present practice is to disposal of wastes by land filling or uncontrolled dumping at the disposal yard the proposed disposal system has been revised synchronizing with the storage and primary collection and taking into account municipal solid waste rules 2000. Land filling shall be restricted to non biodegradable, inert waste and other waste those are not suitable either for recycling or for biological processing land filling shall be carried out for residual of waste reducing policies as well as preprocessing rejects from waste processing facilities. Land fill site shall need the specification as given in schedule 3 of municipal solid waste rules 2000.

The decision to implement any particular technology needs to be based on its techno economic viability, sustainability, as well as environmental implications. The key factors are:

- The origin and the quality of municipal solid waste
- The quantity of waste generated
- Market for final products compost/power
- Commercial fertilizer prices prevailing
- Land price capital and labor cost
- Capabilities and experience of the technology provided.

It needs to be ensured that the proposed facility should fully comply with the environmental regulations laid down in the municipal solid waste rules 2000 issued by ministry of environment and forests, New Delhi.

When the above factors are applied for Mahabubnagar municipality it is recommended to have composting process as option. Composting process

- Technology is techno commercially available
- Technology meets the regularity requirements and is socially acceptable with minimum impacts to the environment and citizens.
- Quantity of waste is less than 150 tons per day making composting operationally feasible.
- Sufficient land availability for establishing of facilities and all other related infrastructure.

## III. CASE STUDY – SOILD WASTE MANAGEMENT in MAHABOBNAGAR MUNICIPALITY:

Mahabubnagar is situated towards the southwest at a distance of 100 km from Hyderabad city. Mahabubnagar is located at 16.73°N, 77.98°E and at an elevation of 493 mts, and spread over an area of 18472 sq.kms. Population of Mahabubnagar is above 139534 as per 2001 census, with 22763 numbers of households. The climate of Mahabubnagar is hot and humid, tropical, summer temperature is 32- 42 centigrade, winter temperature 10° - 32° and with annual rain fall of 355 mm. Mahabubnagar generates about 70 metric tons of waste every day from households, shops and workshops offices and institutions etc..., waste is directly transferred from the primary collection tool into the transportation vehicles the available infrastructure with the municipality 70 tricycles, 10 tractors, hand carts 100. The municipality transporting all collected solid waste to the 25 acre existing dumping yard. The quantity of waste generation is estimated and considered by using the secondary sources and primary survey results.

As per municipal records quantity of waste generation in Mahabubnagar is about 70 MTPD. The sources of the waste contributing to the total tonnage is given in the

**Table: II**

TOTAL QUANTITY OF WASTE GENERATED			
SNo	Source	Total waste TPD	% of Waste
1	Domestic house holds	36.47	52.1
2	Commercial establishment	9.32	13.32
3	Marriage and function halls	2.95	4.21
4	Hotels and lodges	2.84	4.06
5	Markets	2.21	3.16
6	Schools and institutions	1.45	2.07
7	Street sweeping and drain cleaning	13.44	19.2
8	Hospital waste	0.87	1.24
9	construction	0.45	0.64
	Total waste generated	70	100

From the above table it is seen that average per capita generation is about 0.36 kg per capita per day. Apart from the above waste generation was also assessed based on the capacity of each vehicle and the number of trips made in a day to the dumping site. The necessary details were collected and the waste quantity reaching the dumping yard is found to be 45-53 tons per day which translates to about 0.8-0.36 kg per capita per day. However the quantity of waste within the dumping site is about 70-90% of the generated quantity thus per capita generation is 0.36 kg per day.

But the national average per capita generation of waste semi-urban tons is around 0.30 to 0.35 kg per capita per day. Therefore per capita generation for all future projections for Mahabubnagar is taken as 0.36 kg per capita per day. Hence the waste generation for Mahabubnagar town can be taken as (146972\*0.36) is 52.9 TPD say 53 TPD.

Composition of Waste :-

Waste compositions are addressed in the form physical as well as chemical parameters. The following sections give information on both physical and chemical characteristics of waste. Men technique

Physical composition:-

The information on the quantity of waste generated and its composition are the basic needs for the planning of a solid waste management system. Quantity and characteristics of solid waste generated varies with income, socioeconomic conditions, social developments and cultural practices. The characteristic and quantity of waste generated based on the income pattern is presented in following table.

It is noticed that in high income countries the waste generated is more compared to that of low income countries whereas the density of waste is low from high income countries and high in low income countries indicating that more volumes are generated in high income as compared to low income.

**Table: III**

PHYSICAL COMPOSITION OF WASTE OF MAHABUBNAGAR	ITEM WISE GENERATION
ORGANIC WASTE: Comprising of leaves, Fruits, Vegetables, Food Waste, Coal, Fine organic matter, Hay and stray etc...	54.1
RECYCLABLES: Comprising of Rubber and Leather, Plastics, Rags, Paper, Wooden matter, Coconuts, Bones, Straw Fibers’.	12.2
INERT MATTER: Comprising of Ash, Earthen Ware (POTS), Stones and Bricks, Metals, Glass.	33.7
TOTAL	100.00

Chemical Characterisation Of Waste:-

Chemical characteristic considered for municipal waste are mainly moisture, nitrogen, phosphorus, potassium, C/N ratio etc...

Municipal Solid Waste sample was collected from the dump yard for Mahabubnagar town and analyzed for various chemical characteristics. The results are shown in the below table

**Table: IV**

Sl No	ITEM	UNIT	RESULT
1	pH(5% solution)	-	7.12
2	EC(5% solution)	103	415
3	Total Waste Soluble	Mg/gm	21.1
4	Moisture Content	%	40.23
5	Total organic Carbon	%	15.32
6	C/N Ratio	-	1:25
7	Calorific Value	Cal/gm	1241
8	Total Phosphorus	%	0.42
9	Total potassium As K	Mg/gm	3.64
10	Total nitrogen As N	%	0.62
11	Arsenic As As <sub>2</sub> O <sub>3</sub>	Mg/kg	<2
12	Cadmium As Cd	Mg/kg	<0.5
13	Chromium As Cr	Mg/kg	<5
14	Nickel As Ni	Mg/kg	22.3
15	Lead As Pb	Mg/kg	14.10
16	Zinc As Zn	Mg/kg	111.2
17	Copper As Cu	Mg/kg	<5
18	Iron As Fe	Mg/kg	4364

#### **IV. EXISTING WASTE MANAGEMENT PRACTICES – MAHABOOB NAGAR:**

The wastage is stored and transferred from primary collection tool into the transport vehicle. 80% of the population stores the waste at the source on the street open spaces and drains. 80% efficiency has been achieved for implementing the system of segregate of recyclable waste at the source. 80% of households, shops, establishments segregate the waste at the source. The ward wise storage depot details are given below:

Name of the ward- 38

No. of open storage sites- 250

No. of Masonry Bins – Nil

No. of round concrete pipe lines – 250

No. of covered metal containers - 12

Segregation of Recyclable Waste :-With 80% efficiency system of segregation of recyclable waste at the source is done. No special efforts are made by the municipality to educate the people to segregate recyclable waste. Traditional, segregation of recyclable waste is partially practiced by households / commercial establishment.

However the recyclable material is still disposed off by the residents along with domestic waste in a mixed form. This waste finds its way on the streets, in the drains, etc., Recyclable waste is, therefore, generally found mixed with garbage in the domestic bins, on the streets in the municipal bins and at the dumpsites.

Primary collection:-There are 23,000 household, 10,505 commercial and 100 industrial establishments and 75 institutional buildings in the town. System of primary collection of waste from the doorstep has been introduced in 80% households and establishments. The population covered for door to door collection is one lakh. The system of waste collection adopted in the city for collection of household waste, commercial waste, market waste, hotel waste, bio medical waste, construction waste is Two Bin System, Tricycle, Tractors.

Frequency of street sweeping:-The frequency and percentage of cleaning the roads and streets are represented in table:

**Table: V**

Status of Cleaning	% of cleaning the roads & streets
<b>Daily</b>	<b>90%</b>
<b>Alternate Day</b>	<b>5%</b>
<b>Twice a week</b>	<b>5%</b>
<b>Once a week</b>	-
<b>Occasionally</b>	-

The frequency of cleaning bins is done almost every day (90%) and sometimes alternate day (10%). The duty of street sweepers is 8 hrs / day. The work norms adopted by the street sweepers are area wise, ward wise and with assisted trained jawans. Street sweeping is done every day throughout the year including Sundays as well as public holidays. The minimum distance the sweepers had to walk with hand cart to unload the waste storage depot is 50m. The major tools given to safai karmacharis for street sweeping nala cleaning etc., are sped, baskets, axes, crow bar, drain cleaning sped, wheel borrows, shorthanded brooms.

Transportation of Waste :-The number of handcarts is 100 and tricycles are 70 with the solid Waste Management Department. 50% of the sweeper is provided with handcarts and 505 are provided with iron baskets. The quantity of waste transported is measured by visual estimate. The details of transportation and employment are listed below table.

**Table: I**

Types of vehicles	Tractors
No. of vehicles	10 Nos.
No. of Drivers	10 Nos.
No. of shifts in which transportation activity	5 trins hv each tractor
No. of trips made by each type of vehicle in one shift	

The average distance the vehicle has to travel to reach the processing/disposal site is 5 kms. The transportation of the waste is done every day including public holidays and Sunday also. The bio medical waste, hotel waste; construction waste is transported using tricycle. The quantity of waste transported is 35 tones in each shift.

Processing and disposal of waste :-No processing of the waste is being done by the Municipality, the area of land fill site is 5 acres and it will long up to 10 years. The dumping of the waste is done by the tractor and tricycles.

Existing Dump site Details :-In Mahabubnagar, the municipality has introduced door to door collection scheme of solid waste through tricycles. There are sufficient number of tricycles with a capacity of 0.5 Tone each were provided in 38 wards, where as the street and road heap collections were transported through wheel barrows of capacity 0.05 toned disposed primarily in the respective concrete collection and dumper bins. The concrete bins with a capacity of 2 tone and dumper placers with a capacity of 2 tonne were placed at the notified areas of Mahabubnagar Municipality.

The waste is being lifted by the municipal tractors with a capacity of 2 tone, dumper placers with a capacity of 2 tone, tipper with a capacity of 4 tone, and transported all collective solid waste at the selected dumping yard situated at survey No. 921, Koyalkonda x Road. Which are around 25 acres and 5 km away from the town.

**Integrated Waste Management System:** For designing any waste management facilities the points to be noted are the waste quantities generated, design period, waste quantity to be taken for the design period.

**Waste qualification and characterization:-**

Waste quantities depend on the population. The total waste generation from the municipality is measured or estimated based on the population of the town and the per capita waste generation. The future waste generation of the town is also been predicted.

**Table: VII**

Population Projections And Waste Quantities

Name Of The Town	Year	Population	Per Capita Kg/C/D	Waste Quantity TPD
Mahabubnagar	2009	146972	0.36	53
	2019	159470	0.4	64
	2029	171752	0.43	74

**Design Period: -** Processing facilities are designed for the period of 10 years as the life of the processing machinery is generally 10 years. Landfill facility is designed for 30 years period which would be constructed in phases. Landfill is designed for 30 year period.

**Table: VIII**

Waste Generation And Composition

Name Of The Town	Total Waste Generation	Organic Matter	Recyclables	Inert Materials
	TPD	TPD	TPD	TPD
Mahabubnagar	65	35	8	22

**Design of Processing Plant –**

**Compost Plant:-**Composting is the preferred option of processing for Mahabubnagar. Composting is a process of microbial degradation where organic matter is broken down by a succession of organisms in a warm, moist aerobic environment (controlled condition). Composting is form of recycling. Like other recycling effort, the composting of municipal solid waste that must help decrease the amount of solid waste that must be sent to a landfill thereby reducing disposal costs. Composting also yields a valuable product that can be used by the farmers, landscapers, horticulturists, government agencies and property owners as a soil amendment or mulch. The compost product improves the condition of soil, reduces erosion and helps suppress plant diseases.

Composting is an age old practice and the word compost is as old as agriculture itself. The solid wastes of plant and animal origin are utilized for conservation of carbon and mineralization.

It is the decomposition of organic matter by micro-organisms in warm, moist, aerobic and anaerobic environment. The compost made out of urban heterogeneous waste is found to be of higher nutrient value as compared to the compost made out of cow dung and agro-waste. Composting of municipal solid waste is, therefore the most single and cost effective technology for treating the organic fraction of the municipal solid waste. Main advantages of composting include improvement in soil texture and augmenting of micronutrient deficiencies. It also increases moisture holding capacity of the soil and helps in maintaining soil health with a concept of recycling nutrients to the soil. The composting does not require large capital investment, compared to other waste treatment options. At operation level, segregation of municipal solid waste is most important to avoid any toxic heavy metals present in the waste. The compost made from the local municipal solid waste can be marketed near the compost site itself to minimize transportation cost. There are many small and large composting projects in operation in India and

The designing capacities range from 100 to 700 TPD in different locations. Many of the composting facilities are being managed by the private sector through contract arrangements with the municipal authorities. In view of above advantages composting facility is proposed.

**Windrow Platform:-**The concrete yard is an essential infrastructure for preventing contamination of surface/underground water and nearby water bodies. In the instant case the concrete yard is designing in such a way that the fresh garbage received during the first 30 days is decomposed so that the volume and weight of the organic matter is considerably reduced. Inactivation is attained and the stability of organic matter is expected after 30 days.

**Processing Equipments:-** This is an area where the entire waste received is turned at regular intervals. Waste is shifted for feeding to the machinery. Rejects are pushed and the finished materials is also moved to bragg area. The front end pay loaders are essential for above activities.

As there will be a lot of dust and moisture during waste treatment process. These equipments require frequent and constant maintenance and therefore care is taken to provide adequate number of equipments includes 10 wheel tippers for crisscross movement of the waste / manure inside the treatment area and also to deliver finished materials to the required place within the primary marketing zone.

Designing of the concrete yard, processing machineries and equipments have been done in order to ensure treatment of the waste on a day to day basis. In a composting industry waste should not be made to accumulate as it gives out pollution and the cost of holding will also be heavy.

Following schedule will be adopted for turning of the windrows.: 1<sup>st</sup> turning 5<sup>th</sup> day of windrow formation, 2<sup>nd</sup> turning 6<sup>th</sup> day after 1<sup>st</sup> turning 3<sup>rd</sup> turning 6<sup>th</sup> day after 2<sup>nd</sup> turning and screening 6<sup>th</sup> day after 3<sup>rd</sup> turning.

**Windrow turning mechanization and windrow formation:-**Municipal solid waste received on each day will be formed into spate windrow everyday. Incoming vehicles will move only on the outer pathway. They will unload the material in the area designated for the purpose. Soon on unloading biological inoculums will be sprayed on the heap. Hydraulic excavators will lift the material and form the windrow. Outer row will accommodate 6 windrows.

- Daily in the flow of garbage X MT
- Bulk Density of garbage Y
- Volume of garbage received daily X/Y CUM

Cross sectional area of windrow:-  $(A+B)H/2$  SQ.M

Where A = Base width; B = top width; H = Height, all meters,

.Length of windrow in meters, Total volume in cubic meters / cross sectional area in sq meters. An appropriate system is envisaged for continuous draining of leachate generated so that aerobic conditions are maintained inside the windrow for speedy composting.

**Processing section: -** Processing section is divided into three modules namely Preparation section, Secondary section, and Packing section.

**Preparatory section: -** In the preparatory section, the digested municipal solid waste will be passed through 75mm and 30mm trammel screens. The output will be stored in curing shed. The average retention period in the curing shed for 30 days. This will help in improving the quality of the end product. Further it will also increase recovery of composting and reduce rejects for land filling. The curing shed will have a capacity to accommodate 60 days of production.

**Secondary section:-** In the secondary screening section, the cured crude product will further passed through 16mm and 4mm screens. The output will be further passed through specific gravity separator to remove sand and other heavy impurities.

Packing section: -The compost so obtained will be stored in the area and it will be enriched with the nitrogen fixing and phosphorous stabilizing bacteria packed as per demand in the packing section. By splitting the screening operations into modules with adequate intermediate storage capacities, any breakdown of operations in one section will not affect operations in the other section or in other words each section can operate independently without creating bottlenecks. Capacities of each section is designed in such a way that backlogs created due to shut down or break down can be clearly subsequently overcome.

**Table: IX  
Compost Plant Area Requirements**

Sl. No	DESCRIPTION	UNIT	NOS	LENGTH (M)	BREADTH (M)	TOTAL AREA(Sqm)
1	Open windrow platform	Sqm	1	70	30	2129.40
2	Preparatory Section	Sqm	1	10	10	300.00
3	Rejects Section	Sqm	1	4	4	120.00
4	Curing Shed	Sqm	1	20	20	919.29
5	Storage Godown	Sqm	1	10	10	204.29
	Total					3672.98

Proposed Systems for Disposal of inerts / Rejects Arising From Processing Operations:-The Municipal Solid Waste rules 2000 laid down the criteria for disposal of waste as under. Land filling shall also be restricted to non-biodegradable, inert wastes and wastes those are not suitable either for recycling or for biological processing. Land filling shall also be carried out for residues of waste processing facilities as well as pre-processing rejects from the waste processing facilities. Land fillings of mixed waste shall be avoided unless the same is found unsuitable for filling of mixed waste shall be avoided unless the same is found unsuitable for waste processing. Under unavoidable circumstances or till installation of alternate facilities, land filling shall be done following proper norms. Landfill sites shall meet the specifications as given in schedule III of the Municipal Solid Waste rules.

Land Sizing:-The volume of waste to be dumped in the landfill is worked out to 22 Metric Tons per day and the area required for 30 years is 3.00 Hectares.

Landfill design: - Main aspects covering the Landfill Design and Construction are: To minimize the possibility of contaminating surface and ground water, to have control over gaseous emissions and to minimize resource productivity

As suggested by MOEF guidelines a composite liner of two barriers made of different materials, placed in immediate contact with each other provides a beneficial combined effect of both the barriers. The liner system suggested by MOEF is a geo membrane layer over the clay or amended soil barrier. A drainage layer and leachate collection system is placed over the composite liner system.

The effectiveness of the barrier layer basically depends on the hydraulic conductivity of the clay/amended soil layer and the density of the geo membrane against puncture. The clay/amended soil liner is effective only if it is compacted properly and geo membrane liner is effective only if it has the density or mass per unit area (minimum thickness is specified) is sufficient enough against puncture.

Starting from the bottom of the natural ground level, the following layer configuration are proposed for the bottom of the landfill.

**Table: XI Bottom Liner System**

LAYER NO	MATERIAL DESCRIPTION	THICKNESS
Layer 1	Barrier Soil Layer Comprising Of Clay or Amended Soil With Permeability Coefficient Less Than $1 \times 10^{-7}$ cm/sec	900mm
Layer 2	High Density Polyethylene	1.5mm
Layer3	Soil Protection Layer	100mm
Layer 4	Drainage layer	300mm

Assessment of Leachate Quality: - Leachate refers to the liquid that has passed through or emerged from Solid Waste and contains dissolved materials removed from the solid waste. The Leachate generation is primarily a function of precipitation and is directly proportional to rainfall intensity and surface area. Leachate is basically generated from the active landfill area and after closure of land fill site.

Leachate quantity can be estimated for landfill and the leachate network is envisaged in such a way that leachate from the processing facility and landfill will be conveyed to centralized treatment plant to be treated to meet disposal standards.

Formula:-

$$I = P - PCR / O - AET +/- S$$

Where,

I = Rate of Infiltration

P = Precipitation

PCR / O = Coefficient of Runoff

AET = Actual Evapo - Transpiration

for Mahabubnagar is 72 KLD Soil moisture Content Retention

S = Capacity

Empirically,

For Capped portion of landfill:  $I = 0.01 P$

For Uncapped portion of landfill:  $I = 0.07 P$

Landfill with temporary cover:  $I = 0.3 P$

Using the above formula leachate quantity assessed

Leachate Treatment – Process Description :

The treatment process (physical and biological processes) should be adopted in such a way that it should meet the disposal standards. The treatment process suggested is solar evaporation ponds / effluent treatment plant.

**Table: XII Charecterisics of treated leachate**

PARA METER	INLAND SURFACE WATER	PUBLIC SEWERS	LAND IRRIGATION	FOR MARINE COSTAL AREAS
<b>pH</b>	5.5 to 9.0	5.5 to 9.0	5.5 to 9.0	5.5 to 9.0
<b>TSS( mg/l)</b>	100	600	200	Floatable solids max 30mm. Settle able solids max 850microns
<b>BOD (5 days @ 20° C) mg/l</b>	30	350	350	100
<b>COD mg/l</b>	250	-----	-----	250

Landfill gas collection and management system :

As only rejects after processing the municipal waste is proposed to be send landfill is categorized as inert and generated would be very minimal or negligible the quantity of gas generated from the landfill can be estimated with the help of method suggested in CPHEEO manual  $V = CWP/ 100 \text{ m}^3/\text{year}$ .

Design of layers each with a specific function he surface cover system to enhance surface drainage, minimize infiltration support vegetation and control and release of landfill gases. The landfill cover to be adopted depends on the gas management system landfill cover and sequence of its laying: Final landfill cover is usually compose of several and as per the recommendations made by MoEF and CPHEEO.

Top cover layer of 450 mm thick with 300mm thick top soil and 150mm of vegetation supportive soil, drainage layer with 150mm thick clay liner with 600 mm thick and 200mm thick gas collection.

Details of Machinery: Different machinery proposed for the compost plant.

Primary screening:

- Conveyer -3nos
  - Trammel – 35mm screen
  - Trammel - 16mm screen
  - Hydraulic Power Pack – 2 nos
  - Hydraulic piping – 1 lot
  - Electrical Control Panel – 1lot



Secondary Screening :

- Conveyer – 3nos
- Trammel – 4mm screen
- Bucket Elevator – 1 no
- Gravity Separator – 1 no
- Dust collector – 1 Lot
- Packing System – 1 Lot
- Hydraulic Power Pack – 1 No
- Hydraulic Piping – 1 No

Mobile Equipment

- Excavators
- Front End Loaders
- Tippers

Methodology for remediation:

The waste remediations methods are like mining capping are closure etc. For the present conditions, capping of the existing dump site by relocating the dump to one part of the site is proposed and will be further closed according to Municipal Solid Waste Rules 2000. The land after remediation will be used and if it is not sufficient for new processing and disposal facility, new and adjacent to the existing dump site can be acquired for feasibility of operations and handling of new waste.

According to the area requirements of waste lying in the site at different locations to be estimated and assessed for deployment of vehicles for the relocation of the waste to the area earmarked for closure activity will be done depending on the feasibility at that location.

Drainage of surface water runoff Surface water runoff is a significant component in a landfill design and shall be clearly designed. The design includes a garland drainage system all around the landfill which shall be lined and shall be connected to a storm water outlet.

Surface water and Drainage Control Systems

Artificial and natural features at the landfill site control surface water and ground water when integrated, the artificial and natural features must be effective in controlling runoff of surface waters as well as preventing groundwater from penetrating the landfill liner. When the landfill is closed, the drainage control system must be designed to function for the long term use of the site. Rainfall must be used removed from the final cover surface without soil or excessive water infiltration. The greatest risk to the site from pending of surface waters in areas of land subsidence. The features included in the design of drainage control facilities

- 1) Collection and routing of surface waters off the landfill surface in the shortest possible distance
  - 2) Selection of channel and drainage ways that will carry waters at adequate velocities to avoid deposition,
  - 3) use of sufficient surface slopes to maximize the removal of surface runoff and at the same time minimize surface scour and
  - 4) Material specifications for the drainage features that allow repair and replacement as the landfill settles.
- Access road of 5m width shall be provided around the site along with sufficient green belt.

**Conclusion and Recommendations**

To devise any sound waste management systems it is imperative to understand the current practices and scenario of waste management. The physical composition of municipal solid waste is normally presented as organic, recyclables and inert matter. Chemical characteristic considered for municipal waste are mainly moisture, nitrogen, phosphorous, C/N ration etc, Municipal Solid Waste sample was collected from the dump yard of Mahaboobnagar town and analyzed for various chemical characteristics. Since the calorific value of the Municipal Solid Waste is very low 1241Cal/gm, the best process for Municipal Solid Waste management is preferred as composting. The waste projection for 2029 is worked out based on the current generation. Based on the quantities compost processing plant is designed. For the inert resulted after the processing landfill is also designed and leachate system is also proposed to comply Municipal Solid Waste Guidelines as per CPHEEO.

**A. Conclusion:**

- Environment Protection Act concludes aerobic composting does not contribute to CO<sub>2</sub>, CH<sub>4</sub> or N<sub>2</sub>O emissions, the main contributors to greenhouse gas response and global warming.
- Any emissions from aerobic composting are considered part of the natural carbon cycle.
- Proper aerobic composting eliminates methane production
- Aerobic compost can be used as a landfill cover to reduce and eliminate methane emissions and odour as a result
- Aerobic composting appears to be the safest way of converting organic waste streams into a stable value added product
- Carbon is essential for soil stability and fertility
- Aerobic compost a sink helping reduce emissions in the atmosphere by sequestering the carbon in the soil.
- Converting organic waste into aerobic compost provides us with the potential to existing vegetation, allowing for more respiration in our atmosphere, which results in a reduction in CO<sub>2</sub> levels.

**B. Recommendation:**

In the present study of composting plan for Mahabubnagar Municipality the Green House Gases focused are Methane and Carbon dioxide.

Composting is a financially viable option for municipalities with solid waste range of 150MT and below. It helps reducing Green House gas emissions thus complying to the projects under Clean Development Mechanism.

**REFERENCES:**

- [1.] Manual on Solid Waste Management and Handling Rules
- [2.] Air Pollution C.S. Rao
- [3.] Waste Water Management MetCalf&Eddy

**BIOGRAPHY**

**First Author:**



**Dr. C.Sarala**, Associate Professor, Centre for Water Resources, Institute of Science and Technology, Jawaharlal Nehru Technological University Hyderabad. Her research interest includes surface water resources analysis and environment related problems.

# Survey Paper on Creation Of dynamic query Form for mining highly optimized transactional databases

Jayashri M. Jambukar,

PG student, Amrutvahini college of Engineering, Sangamner.

## ABSTRACT

In New scientific databases and web databases maintain huge and heterogeneous data. These concrete world databases include over so many relations and attributes. Historic predefined query forms are not able to answer different ad-hoc queries from users on those databases. This paper proposes Dynamic Query form, a curious database query form interface, which is able to dynamically create query forms. The significance of DQF is to capture a user's choice and classify query form components, support him/her to make conclusion. The creation of query form is a repetitive process and is conducted by the users. In each repetition, the system automatically creates classification lists of form components and the user then adds the desired form components into the query form. The classification of form components is based on the captured user choice. A user may also fill up the query form and deliver queries to view the query output at each step. Thus, a query form could be dynamically refined till the user answer with the query output. A probabilistic model is developed for estimating the excellence of a query form in DQF. I have studied evaluation and user study certify the effectiveness and efficiency of the system.

**Keywords:** Form creation, Query Form, User Interaction.

## I. INTRODUCTION:

Query form is one of the most extensively used user interfaces for querying databases to access information. Historic query forms are configured and predefined by developers or Database Administrator in different information management systems. With the fast development of web information and scientific databases, new databases become very huge and difficult. In natural sciences, like genomics and diseases, the databases have number of entities for chemical and/or biological data resources. Different types of web databases, like Freebase and DBpedia, have thousands of structured web entities. Therefore, it is difficult to design a set of static query forms to answer various ad-hoc database queries on those difficult and complex databases.

Many existing database management and development tools, like EasyQuery, Cold Fusion, SAP and Microsoft Access, provide various mechanisms to let users generate customized queries on databases. But, the customized queries generation totally depends on users' manual editing's. If a user is not familiar with the database schema in advance, those hundreds or thousands of data attributes will confuse him or her.

### 1.1 INTERACTION BETWEEN USERS & DQF

- A. Query Form Enrichment
  - 1) Dynamic Query Form (DQF) recommends a ranked list of query form components to the user.
  - 2) The user has to select the desired form components into the current query form.
- B. Query Execution
  - 1) The user fills out the current query form and submits a query.
  - 2) DQF will execute the query and the results are shown.
  - 3) The feedback about the query results is provided by user.

## II. SYSTEM ARCHITECTURE

The system is proposed to have the following modules along with functional requirements.

- A. Query Form Enhancement
- B. Query Execution
- C. Customized Query Form
- D. Database Query Recommendation

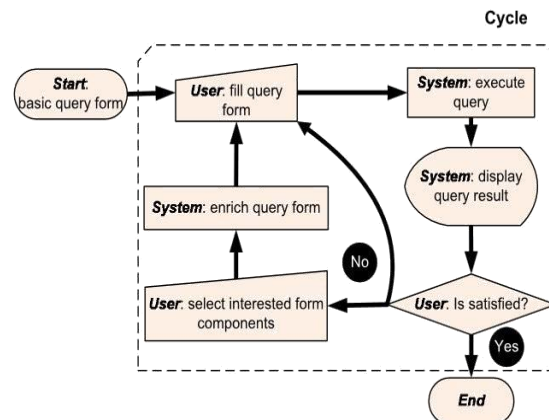


Fig 1: System Architecture

### A. Query Form Enhancement

- 1) Dynamic Query Form endorses a ranked list of query form components for the user.
- 2) The user has to select the preferred form components into the current query form.

### B. Query execution

- 1) The user has to fill out the current query form and submits the query.
- 2) DQF performs the query and displays the results.
- 3) The user offers the feedback on the query results.

### C. Customized Query Form

These provide visual interfaces for developers to generate or customize query forms. The issue of those tools is that, they are for the professional programmer who is aware with their databases, but not for the end-users. It suggests a system which permits end-users to customize the existing query form at run time. But, the end-user may not be familiar with the database. If the database schema is very huge, it is hard for them to search specific database entities and attributes and to generate desired query forms.

### D. Database Query Recommendation

Current studies introduce shared method to recommend database query components for database research. They consider SQL queries as elements in the collaborative filtering strategy, and proposes similar queries to relevant users.

## III. ALGORITHM

Below algorithm shows the algorithm of the One-Query's query creation. The function createQuery is to create the database query centered on the given group of projection attributes  $A_{one}$  with selection expression  $\sigma_{one}$ .

Data:  $Q = \{Q1, Q2, \dots\}$  This is the set of earlier queries executed on  $F_i$ .

Result:  $Q_{one}$  is the query of One-Query begin

$\sigma_{one} \leftarrow 0$  for  $Q \in Q$  do

$\sigma_{one} \leftarrow \sigma_{one} \vee \sigma_Q$

$A_{one} \leftarrow A_{Fi} \cup A_r(F_i)$

$Q_{one} \leftarrow \text{CreateQuery}(A_{one}, \sigma_{one})$

When the system gets the result of the query  $Q_{one}$  from database engine, it requests the second algorithm of One-Query to search<sub>2</sub> best query condition.

### 3.1 Query Form Interface

#### a) Query Form-

Every query form resembles to an SQL query template. Definition 1: A query form  $F$  is defined as a tuple  $(AF, RF, \sigma F, \bowtie (R))$ , this signifies a database query template like in below:

$F = (\text{SELECT } A1, A2, \dots, A_k$   
 $\text{FROM } \bowtie (RF) \text{ WHERE } \sigma F),$

where  $AF = \{A1, A2, \dots, A_k\}$  are  $k$  attributes of projection,  $k > 0$ .  $RF = \{R1, R2, \dots, R_n\}$

which is the group of  $n$  relations included in this query,  $n > 0$ . Every attribute in  $AF$  will belong to one relation in  $RF$ .  $\sigma F$  is the conjunction of expressions for selections on relations in  $RF$ .  $\bowtie (RF)$  is a join function to create a conjunction of expressions for joining relations of the  $RF$ . In user interface of a query form  $F$ ,  $AF$  is the group of columns of result table.  $\sigma F$  is group of input

components to fill for users. Query forms permit users to fill parameters to create various queries.  $RF$  and  $\bowtie (RF)$  are not visible in end user interfaces, which are generally created by system as per the database schema. For query form  $F$ ,  $\bowtie (RF)$  is automatically constructed as per foreign keys among relations in  $RF$ .

In the meantime,  $RF$  is determined by  $AF$  and  $\sigma F$ .  $RF$  is union group of relations which has at least one attribute of  $AF$  or  $\sigma F$ . So as, the components of query form  $F$  are in actual determined by  $AF$  and  $\sigma F$ . As mentioned, only  $AF$  and  $\sigma F$  are visible to user in user interfaces. We focus on projection & Selection components of a query form. Ad-hoc join is not touched by our dynamic

query form because join is not a part of query form and is not visible for the users. As for "Aggregation" and "Order by" in SQL, there are limited choices for users. Like, "Aggregation" can be MIN, MAX, AVG, and so on; and "Order by" can be "increasing order" & "decreasing order". Our dynamic query form can be easily enhanced to include those options by implementing them as dropdown boxes in user interface of query form.

#### b) Query results-

To conclude if a query form is required or not, a user doesn't have time to go over each data instance in query results. Also, many database queries results a large amount of data instances. We only output a compressed output table to display a highlevel view of the query results. Every instance in compressed table signifies a group of actual data instances. Next, user can click through desired clusters to view detailed data instances. Below figure shows user action flow. The compressed view of query results will be proposed. There are many clustering algorithms for creating compressed view efficiently. For our implementation, we select incremental data clustering framework because of efficiency issue. Different clustering methods are preferable to different data types. Here, clustering is just to give a better view of query results for users. The system programmer can choose a various clustering algorithm if required.

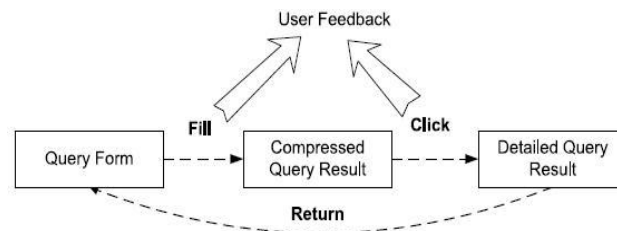


Fig 2: User Actions

**c) Ranking Metrics**

Query forms are developed to return user’s anticipated result. There are two traditional measures to estimate quality of the query outputs: **precision** and **recall**. Query forms are able to generate different queries by various inputs, and various queries can output different query results and obtain different **precisions** and **recalls**, so we are using *desired precision* and *expected recall* to calculate the expected performance of the query form. *Expected precision* is the expected proportion of query results which are interested by user. *Expected recall* is expected proportion of user interested data instances which are returned by current query form. User interest is anticipated based on user’s click through on query results showed by the query form.

Like, in case some data instances are clicked by user, those data instances should have vital user interests. So, query form components which can capture those data instances should be ranked at high than remaining components. Afterwards we introduce some notations and then define desired precision and recall.

ID	C1	C2	C3	C4	C5
I1	A1	B1	C1	30	1
I2	A2	B2	C2	30	100
I3	A3	B2	C3	40	99
I4	A4	B1	C4	30	1
I5	A5	B3	C4	20	2

Table 1:- Example of data table

Like take a query form with one relational data table as shown in the Table. There are 5 data instances in the table,  $D = \{I1, I2, \dots, I5\}$  which has 5 data attributes  $A = \{C1, C2, C3, C4, C5\} = 5$ .

Query form executes a query  $Q$  as “SELECT C2, C5 FROM D WHERE C2 = B1 OR C2 = B2”. The query result is  $DQ = \{I1, I2, I3, I4\}$  which are projected on C2 and C5.

Thus  $P(\sigma_{Fi} d)$  has 1 for I1 to I4 and has zero for I3. Instance I1 and I4 has same Projected values.

So we use I1 to represent both of them and  $P(I1C2; C5) = 2/5$ .

**IV. STATIC VS. DYNAMIC QUERY FORMS**

When a query task is covered by one historical queries, then SQF built on those historical queries can be used to fill that query task <sup>3</sup>But the costs of using SQF and DQF to fulfill those task are different. *Form-Complexity* was proposed in to estimate cost of using a query form. That is sum of the number of selection components, projection components, and Relations.

**V. USABILITY METRICS**

For database query forms, one *action* means a mouse click or a keyboard input of a textbox. *ACmin* is a minimal number of *actions* for a specific querying task. One *function* signifies a provided option for user to use, like a query form or a form component. In case of web page based system, *FNmax* is total number of UI components in web pages explored by users.

Here each page at most contains 5 user interface components. The smaller  $AC_{min}$ ,  $AC$ ,  $FN_{max}$ , and  $FN$ , the better will be the usability. And higher the  $AC_{ratio}$ ,  $FN_{ratio}$ , and  $Success$ , the better will be the usability. There is a trade-off between  $AC_{min}$  and  $FN_{max}$ . The extreme case will be when, we create all possible query forms in one web page, and user only needs to select one query form to complete their query task, so  $AC_{min}$  is 1. However,  $FN_{max}$  should be number of all possible query forms with their components, which can be a large number. On other side, when users have to interact a lot with a system, that system should know better about user's anticipation. In such case, the system would cut down many unwanted functions, so that  $FN_{max}$  will be smaller. But  $AC_{min}$  will be high since there are many of user interactions.

Effectiveness-

Here we compare ranking function of DQF with other two ranking methods: baseline method and other is random method. Baseline method ranks projection and selection attributes in ascending order of their schema distance to current query form. In case of the query condition, it selects the most frequently used condition in training set for that particular attribute. Random method randomly proposes one query form component. Final truth of the query form component ranking is obtained from the query workloads.

Here we use some widely used metrics in Human-Computer Interaction and Software Quality for measuring the usability of a system. These metrics are listed in below Table:

Metric	Definition
$AC_{min}$	The minimal number of <i>action</i> for users
$AC$	The actual number of <i>action</i> performed by users
$AC_{ratio}$	$AC_{min}/AC \times 100.0\%$
$FN_{max}$	The total number of provided UI <i>function</i> for users to choose
$FN$	The number of actual used UI <i>function</i> by the user
$FN_{ratio}$	$FN/FN_{max} \times 100\%$
$Success$	The percentage of users successfully completed a specific task

Table 2:- Usability Metric

## VI. CONCLUSION

I studied dynamic query form generation approach which helps users dynamically generate query forms. The key idea is to use a probabilistic model to rank form components based on user preferences. We capture user preference using both historical queries and run-time feedback such as click through. Experimental results show that the dynamic approach often leads to the higher success rate and simpler query forms compared with a static approach. Ranking of form components also makes it easier for users to customize query form.

## REFERENCES

- [1] DBPedia. <http://DBpedia.org>.
- [2] Easy Query. <http://devtools.korzh.com/eq/dotnet/>
- [3] Freebase. [Http://www.freebase.com](http://www.freebase.com).
- [4] C.C.Aggarwal, J Han, J. Wang, & P. S. Yu. A framework for clustering evolving data streams. In *Proceedings of VLDB*, pages 81–92, Berlin, Germany, Sept 2003.
- [5] S. Agrawal, S. Chaudhuri, G. Das, and A. Gionis. Automated ranking of database query results. In *CIDR*, 2003.
- [6] S. Chaudhuri, G. Das, V. Hristidis, and G. Weikum. Probabilistic information retrieval approach for ranking of database query results. *ACM Trans. Database Syst. (TODS)*, 31(3):1134–1168, 2006.
- [7] G. Das and H. Mannila. Context-based similarity measures for categorical databases. In *Proceedings of PKDD 2000*, pages 201–210, Lyon, France, Sept 2000.

# Image Performance Tuning Using Contrast Enhancement Method with Quality Evaluation

Alphy George<sup>1</sup>, S John Livingston<sup>2</sup>

<sup>1</sup> PG Student, Department of Computer Science and Engineering, Karunya University, Tamilnadu, India

<sup>2</sup> Assistant Professor, Department of Computer Science and Engineering, Karunya University, Tamilnadu, India

## ABSTRACT:

Quality of an image plays a very crucial role in various image processing applications such as, recognition, identification, transmission etc. Therefore identifying the quality of an image is very necessary in such areas. Restoration of the good image from a degraded image will improve the quality of the image. The purpose of this paper is to find the quality value of an image using a new metric after some preprocessing steps. Here one particular type of image distortion taken into account that is contrast change and enhancing the contrast using an additive gamma correction method. After this preprocessing the quality value is found using a new image quality assessment metric called normalized perceptual information distance. For this metric, the main concept used is kolmogorov complexity and normalized information distance.

**Keywords:** Contrast enhancement, Image quality assessment, gamma correction, normalized information distance, kolmogorov complexity.

## I. INTRODUCTION

Image quality has an important role in many image processing applications mainly in the case of recognition. Quality of an image can be degraded because of the presence of various distortions like blur, contrast change, noise, blocking artifacts etc. A low quality or degraded image may not match with the existing image in the database. So image quality measurement is important. In the above case if some preprocessing is done to remove some distortions, then it will increase the image quality. A quality measurement method is the used to measure the image quality.

For measuring the quality, image quality assessment (IQA) algorithms are used. These IQA methods are classified into three, 1) no reference 2) reduced reference and 3) full reference [1]. In no reference method, only distorted image will be available. No information about the original image is given. Because of this, blind calculation of the image quality is possible. In the case of reduced reference method, a portion of the original reference image will be available whereas, in full reference, a full original image will be given. The quality metric can be calculated by comparing distorted image with the original image. In full reference approach, many quality metrics are available for calculating the quality. Structural similarity index measure (SSIM) [1], visual information fidelity (VIF) [2], visual signal to noise ratio (VSNR) [3] etc. is some of the full reference methods.

This paper investigating a full reference image quality assessment metric based on kolmogorov complexity and normalized information distance (NID) [7]. These two concepts are rarely been studied in the field of image processing. As a new method of introducing kolmogorov complexity in image quality assessment, it shows comparable performance with the other existing methods.

One particular type of distortion usually present in images is contrast change. This may occur due to the bad lighting condition or because of the camera problem. In this paper we combined a preprocessing method for image contrast enhancement with the quality metric calculation. There are many methods available for contrast enhancement. Here, contrast enhancement based on gamma correction and a weighting distribution function is used. A normalization function is also applied to the enhanced image to improve the quality of the image.



This paper is organized as follows. In section 2, contrast enhancement method using additive gamma correction (ACG) [4] is explained. This contrast enhancement method is taken as a preprocessing step and embedded to the final image quality assessment framework. In section 3, a full reference image quality measurement based on kolmogorov complexity and normalized information distance [6] is explained. In section 4, the results after implementing this method are given. Finally, in section 5, conclusions are described.

## II. CONTRAST ENHANCEMENT USING ACG

The problem with most of the contrast enhancement methods is their high computational time and cost. The proposed approach is resolving this problem. In ACG method [4], the densely associated pixels in the input dimmed image are smoothed by using a weighting distribution. The flow chart for ACG is given in Fig. 1.

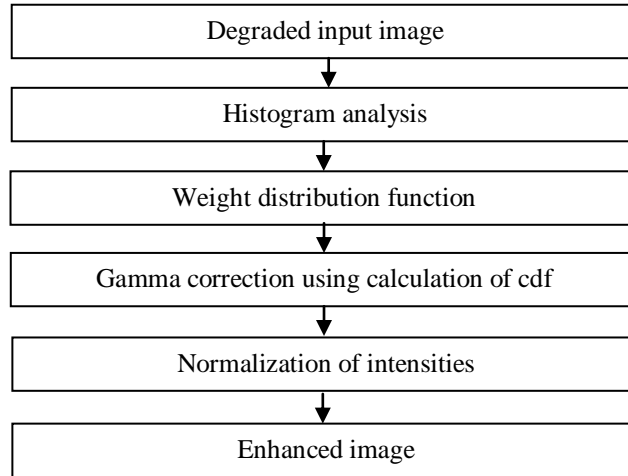


Fig. 1 Flow chart of the ACG method

Firstly, a dimmed image is given as input to this method. Histogram analysis can be done to see the densely oriented pixels. The gamma correction function used here is,

$$T(l) = l_{max}(l/l_{max})^\gamma = l_{max}(l/l_{max})^{1-cdf(l)} \quad (1)$$

where  $l_{max}$  gives the maximum intensity of the dimmed input image.  $T(l)$  represents the transformed intensities of the dimmed image after gamma correction. This gamma correction method includes cdf calculation. For cdf calculation, probability density function (pdf) has to be calculated. A weighting distribution function is using here for pdf calculation. This will modify the histogram and decrease any adverse effects. The weighting distribution function is,

$$pdf_w(l) = pdf_{max} \left( \frac{pdf(l) - pdf_{min}}{pdf_{max} - pdf_{min}} \right)^\beta \quad (2)$$

where  $pdf_{max}$  is the maximum pdf and  $pdf_{min}$  is the minimum pdf.  $\alpha$  is the adjusting parameter. This can be used to find cdf as follows,

$$cdf_w(l) = \sum_{i=0}^{l_{max}} pdf_w(i) / \sum pdf_w \quad (3)$$

sum of  $pdf_w$  is calculated as follows,

$$\sum pdf_w = \sum_{i=0}^{l_{max}} pdf_w(i) \quad (4)$$

This can be applied in Eq. 1 to get the transformed intensities. While implementing this algorithm, first the image is converted into HSV. Here H and S contain the color information. We are preserving the color information and changing only value V. Sometimes the output enhanced image may be appeared as whitish after gamma correction. In order to solve this problem, a normalization of the intensity levels of the output image is used. Equation for that is given in Eq. (2.6).

$$I'_j = ((w \times I_j) + T(I_j))/(w + 1) \tag{5}$$

where  $I_j$  are original intensities of the input image.  $T(I_j)$  are the intensities produced after gamma correction.  $w$  is the non negative weight.  $w$  is selected based on the improvement in the quality.

### III. IMAGE QUALITY ASSESSMENT USING NPID

The problem with most of the contrast enhancement methods is their high computational time and cost. The overall architecture of the proposed approach is given in Fig. 2.

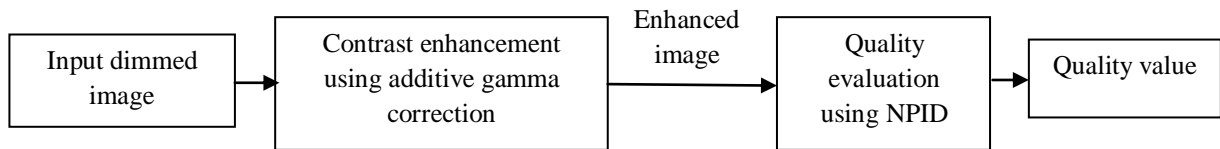


Figure 2. Flow chart for the proposed system

Contrast enhancement method is described in section 2. For quality evaluation using NPID [6], important mathematical concepts used are kolmogorov complexity and normalized information distance (NID). These two concepts are rarely studied in the context of image processing. Kolmogorov complexity [8] is a non computable concept. General equation for NID is,

$$NID(P, Q) = \frac{\max \{K(p|q^*), K(p^*|q)\}}{\max \{K(p), K(q)\}} \tag{6}$$

where  $K(p|q^*)$  is the kolmogorov complexity of two objects  $x$  and  $y$ . For the implementation of NPID metric, first transform the original and the reference image into the wavelet domain. The two images are decomposed into subbands in that domain. Particular type of wavelet transform used here is 5 scale laplacian pyramid transformation [11]. Then kolmogorov complexity can be formulated as,

$$K(p|q) = \sum_{i=1}^n K(p_n|q_n) \tag{7}$$

where  $p_n$  and  $q_n$  are the subbands of reference image and the distorted image. Using this, NID can be calculated as,

$$NID_s(p, q) = \frac{\max \{\sum_{i=1}^n K(p_n|q_n), \sum_{i=1}^n K(q_n|p_n)\}}{\max \{\sum_{i=1}^n K(p_n), \sum_{i=1}^n K(q_n)\}} \tag{8}$$

Because of the non computable nature of kolmogorov complexity, NID is also non computable. To solve this problem, kolmogorov complexity can be replaced with Shannon entropy. Shannon entropy gives how much information contained in an image. And also for finding information content, subband can be modeled using a gaussian scale mixture model [9]. So the final formula for finding NPID image quality metric is,

$$NPIS_i = \frac{I(\overline{M}_i; \overline{N}_i)}{\max \{I(\overline{M}_i; \overline{A}_i), I(\overline{N}_i; \overline{A}_i)\}} \tag{9}$$

Let  $\overline{A}_j = (\overline{A}_{1j}, \dots, \dots, \overline{A}_{Nj})_j$  are the  $N$  elements of subband  $A_j$ , and  $\overline{M}_j, \overline{N}_j$  are defined in the same way. Information content between the subbands can be calculated as [10],

$$I(\overline{A}_j; \overline{M}_j) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K \log_2 \left( 1 + \frac{s_i^2 \lambda_k}{\sigma_n^2} \right) \tag{10}$$

$$I(\overline{A}_j; \overline{N}_j) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K \log_2 \left( 1 + \frac{g^2 s_i^2 \lambda_k}{\sigma_n^2 + \sigma_v^2} \right) \quad (11)$$

$$I(\overline{M}_j; \overline{N}_j) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^K \log_2 \left( \frac{[g^2 s_i^2 \lambda_k + (\sigma_n^2 + \sigma_v^2)](s_i^2 \lambda_k + \sigma_n^2)}{[(\sigma_n^2 + \sigma_v^2)s_i^2 + \sigma_n^2 g^2 s_i^2] \lambda_k + \sigma_n^2 (\sigma_n^2 + \sigma_v^2)} \right) \quad (12)$$

To calculate the information content estimation of the parameters  $g, s, \sigma, \lambda$  are important. These parameters are calculated based on the information from [6].

#### IV. RESULT ANALYSIS

The proposed algorithm is applied on the images in CSIQ (Categorical Image Quality) database [5]. This database is publicly available for download. The database contains 866 distorted images for 30 reference images. Different types of distortions include contrast change, jpeg compression, noise, blur etc. Each distortion is at four to five distortion levels.



Figure 3. Some of the reference images present in CSIQ database



Figure 4. Five levels of contrast change for child\_swimming image in CSIQ database and its corresponding contrast enhanced images.

For the proposed approach, contrast change distortion is taken into account. For each image, five levels of contrast change is there in the database. For each level, contrast enhancement is done, and to the enhanced image, the quality metric is applied. The result obtained after performing the contrast enhancement for some of the images in the CSIQ database is given in Fig. 4.

The quality metric values obtained for the enhanced images after performing the image quality assessment algorithm based on normalized perceptual information distance (NPID) is given in Table 1.

Table 1. Results obtained before and after performing the algorithm to the images in CSIQ database

Image name	1 <sup>st</sup> level		2 <sup>nd</sup> level		3 <sup>rd</sup> level		4 <sup>th</sup> level		5 <sup>th</sup> level	
	before	after	before	after	before	after	before	after	before	after
1600	0.9057	0.7675	0.8963	0.8037	0.8419	0.8454	0.8194	0.8477	0.8194	0.8473
Aerial_city	0.8886	0.7273	0.8488	0.7712	0.7495	0.8261	0.7083	0.8222	0.7083	0.8222
boston	0.8921	0.7043	0.8433	0.755	0.7084	0.8059	0.6277	0.7868	0.6277	0.7868
bridge	0.9115	0.6655	0.8948	0.6812	0.8582	0.7071	0.8398	0.7207	0.8398	0.7207
butter_flower	0.9014	0.7324	0.8653	0.7815	0.7245	0.7909	0.6913	0.7241	0.6913	0.7241
child_swimming	0.8934	0.6295	0.8461	0.6854	0.7107	0.7567	0.646	0.7489	0.646	0.7489
couple	0.9201	0.6138	0.9013	0.6301	0.8568	0.6538	0.8377	0.6607	0.8377	0.6607
family	0.8824	0.7255	0.8476	0.7668	0.7739	0.8239	0.724	0.8275	0.724	0.8275
lake	0.9145	0.7257	0.8861	0.7673	0.8182	0.8253	0.7893	0.8363	0.7893	0.8363
log_seaside	0.9161	0.5775	0.8969	0.5946	0.8523	0.6202	0.835	0.6269	0.835	0.6269
monument	0.9154	0.651	0.8958	0.6729	0.8518	0.7116	0.8378	0.7216	0.8378	0.7216
native_american	0.9069	0.6664	0.8839	0.6863	0.8325	0.7156	0.8098	0.7229	0.8098	0.7229
redwood	0.9032	0.7592	0.8761	0.7904	0.8099	0.8351	0.7869	0.8413	0.7869	0.8413
roping	0.9031	0.7203	0.8765	0.7557	0.8172	0.8049	0.782	0.8147	0.782	0.8147
rushmore	0.9054	0.4891	0.8814	0.507	0.8282	0.5347	0.8064	0.543	0.8064	0.543

Results from the above figure shows that the contrast enhancement function works well if the image quality lies in the range less than 0.8. If the quality value is greater than 0.8 then enhancing the image may decrease the image quality.

## V. CONCLUSIONS

In this paper, we combined the performance tuning of images with quality evaluation. Major two steps performed in the proposed method is contrast enhancement using additive gamma correction and quality metric calculation using normalized perceptual information distance. The algorithm is implemented on publicly available CSIQ database and the results are obtained. If the quality is very poor, the contrast enhancement works well and quality is improved in a good manner. If the image is of good quality, then no change is needed. This work can be extended for finding the quality of video after performance tuning of video frames.

## REFERENCES

- [1]. Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P., "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. Image Process.* 13(4), 600–612, 2011.
- [2]. Sheikh, H.R. and Bovik, A.C., "Image information and visual quality", *IEEE Trans. Image Process.* 15(2), 430–444, (2006)
- [3]. Chandler, D. M. and Hemami, S. S., "Vsnr: a wavelet-based visual signal-to-noise-ratio for natural images", *IEEE Trans. Image Process.* 16, 2284–2298, 2007.
- [4]. Huang S., H., Cheng F., Chiu Y., "Efficient contrast enhancement using adaptive gamma correction with weighting distribution" *IEEE Trans. Image Process.* Vol.22, 2013.
- [5]. Larson, E. C., Chandler, D.M.: *Categorical image quality (CSIQ) database.* (Online) Available <http://vision.okstate.edu/csiq>.
- [6]. Nikwand M. and Wang Z., "Image distortion analysis based on normalized perceptual information distance", Springer, SIViP 7:403-410, 2013.
- [7]. Li, M., Chen, X., Li, X., Ma, B., and Vitanyi, P.M.B., "The similarity metric", *IEEE Trans. Inf. Theory* 50,3250-3264, 2004.
- [8]. Li, M., and Vitanyi, P., "An introduction to kolmogorov complexity and its applications", 2<sup>nd</sup> edn. Springer, Berlin, 1997.
- [9]. Wainwright, M.J., Simoncelli, E.P.: Scale mixtures of gaussians and the statistics of natural images. In: Solla, S.A., Leen, T.K., Müller, K.-R. (eds.) *Advances in Neural Information Processing Systems*, vol. 12, pp. 855–861. MIT Press, Cambridge, 2000.
- [10]. Wang, Z., Li, Q.: Information content weighting for perceptual image quality assessment. *IEEE Trans Image Process.* 20(5), 1185–1198 (2011)
- [11]. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. *IEEE Trans. Commun.* 31, 532–540 (1983)

## **BIOGRAPHIES**



**Alphy George** received her B.Tech (Computer Science and Engineering) degree from SCMS school of Engineering and Technology affiliated by Mahatma Gandhi University in the year 2012. She is pursuing her M. Tech degree in Computer Science and Engineering at Karunya University.



**S John Livingston** received his M.E. degree from karunya university. He is currently working as an Assistant Professor at Karunya University.

# Customized Query Interface Integration using Attribute Constraint Matrices

Sherlin Mary Koshy<sup>1</sup>, Belfin R. V.<sup>2</sup>

<sup>1</sup> PG Student, Department of Computer Science and Engineering, Karunya University, Tamilnadu, India

<sup>2</sup> Assistant Professor, Department of Computer Science and Engineering, Karunya University, Tamilnadu, India

## ABSTRACT:

*Query Interfaces are often the only means by which certain web databases in a domain can be accessed. It has been observed that several different kinds of query interfaces exist for the same domain and they require the user to enter the same values despite the differences in the interfaces. Integrating these query interfaces is therefore essential to allow the user uniform access to all databases in a given domain. Integration of Query Interfaces using Attribute Constraint Matrices was found to be a very efficient technique for integration. This technique however performs integration over all the interfaces in a domain, while integration of only a subset of these may be necessary. This paper presents an algorithm that when applied to the result of integration of all interfaces in a domain produces an integrated attribute constraint matrix that represents only the components of those interfaces that are required to be integrated by the user.*

**Keywords:** Attributes, Attribute Constraint Matrix, Interface Integration, Integrated Matrix, Pruning, Query Interface, User Preference.

## 1. INTRODUCTION

Online web databases contain a lot of useful information and are often accessible only by means of Query Interfaces. Query Interface is a term used to refer to the collection of HTML elements such as text boxes, radio buttons, selection lists etc., that are used to obtain information from the user. Often there are several different query interfaces in existence for a given domain requiring the user to enter the same information multiple times. For example, if a user wants to buy an air ticket online at the cheapest price he would have to visit several web sites and repeatedly enter the same information i.e., departure and arrival cities, date of journey, number of passengers etc., and then compare the price offered by these websites. Hence from the user point of view it would be easier to enter all this information in a single webpage that would appropriately access all these websites and display the results to the user in a single page. The first step to achieve this would be to generate a universal query interface that captures the structure of all the query interfaces of these websites.

There are several techniques in existence that have been summarized in [8] of which the utilization of attribute constraint matrices as defined in [7] has been found to be the most efficient. Attribute constraint matrices [7] are a far easier means of interface representation than that defined by the WISE-Integrator method [1],[2],[3] and also does not require any kind of hypothesizing of the unified interface. The integration technique suggested by [5] requires pair-wise merging which is time consuming when compared with [7]. The merging via clustering aggregation suggested in [4] will cause the attribute set to grow to be very large, while in [7] the attribute set will be restricted since a common representation is taken for the attributes of the same type. Keyword matching technique described in [6] requires additional files to store matched information, which is not required in the case of [7].

The attribute constraint matrix method in [7] first requires an ordered tree representation of the query interface based on the visual layout of the attributes. All leaf elements of the tree correspond to attributes on the query interface, a common representation for the attributes is first determined. For example in the airline domain one website may specify the terms 'departure' and 'arrival' cities while another may define them as 'from' and 'to' cities. Since they both represent the same information the 'departure' and 'from' attributes will both be represented commonly by the letter 'a' for example and the 'arrival' and 'to' attributes will commonly be represented by the letter 'b'. Once these common representations have been determined the ordered attribute trees are created based on the visual layout of the attributes in the interface. From the ordered attribute trees the attribute constraint matrices are generated. These matrices are symmetric matrices having as many rows and columns as there are attributes in the interface. It has been demonstrated in [7] that these matrices can be derived from the ordered trees and vice versa. The values within the attribute constraint matrices consist of depth of the attributes and the distance between the attributes derived from the corresponding ordered tree. These matrices are then suitably expanded and merged as per the operations suggested in [7] to yield the final unified attribute constraint matrix. From this unified attribute constraint matrix the ordered tree representation of the final unified interface can be easily determined.

The performance of this technique [7] was measured in terms of the extent to which the structure of the existing interfaces was maintained in the integrated interface i.e., the extent to which the constraint imposed by each interface was satisfied in the unified interface. There were three constraints described in [7] hierarchical, group and precedence and the satisfaction rate of these constraints were found to be acceptable. Hence we can conclude that the technique of using attribute constraint matrices for generating the unified query interface for all the interfaces in a domain is highly efficient and easy to implement.

It may often be necessary to integrate only a sub set of the interfaces in a domain, for example if only the results of query interfaces within a region is needed, i.e., a user wants to compare the air ticket prices offered by only the travel operators in his country of residence. Or if user wants to view results from websites that offers certain privileges. There are two ways to approach this problem as defined in [9], one is the bottom up approach which involves performing integration from scratch on only the interfaces that are necessary, and the other is the top down approach which involves reducing a global integrated tree representation to contain only the attributes of the interfaces that are preferred. This top down approach called pruning [9] is defined to work on the global integrated tree. The aim of this paper is to perform the same pruning operation on the global unified attribute constraint matrix generated by [7].

Pruning the integrated tree is a far more efficient technique than generating the integrated tree bottom up because the pruning operations become easier with an increase in the number of the interfaces that need to remain. This is because with an increase in the number of interfaces that need to remain there will be fewer attributes that need to be removed from the integrated tree hence fewer operations that need to be performed, this is not the case with the bottom up approach since more attributes will need to be included in the integrating process resulting in an increased amount of time being taken.

The rest of this paper is organized as follows. In section 2, the top down tree pruning technique described in [9] is explained. This serves as the basis of the main work in this paper. In section 3, an algorithm is presented and explained that performs the same pruning operations except that the algorithm takes as input the unified attribute constraint matrix generated by [7] as the input. In section 4, the results after implementing this method are given. Finally, in section 5, the conclusions are described.

## II. TOP DOWN TREE PRUNING

This technique described in [9] works on the integrated tree schema that has been generated using the methods described in [1] and [5]. The essential step is to remove from the integrated schema those fields that are non-existent in the interfaces that are required. Once fields are removed there needs to be further operations that need to be done to refine the final tree.

The suggested technique has three primary steps, Remove, Collapse and Split-Up. The remove step involves removing from the integrated tree all the leaf nodes that do not exist in the interfaces being considered, this step also recursively removes all 'fake' leaves, i.e., those internal nodes that become leaves in the process of removal. The collapse step collapses a child node with its parent if it is the only child, this step is also performed recursively. The final step splits up the nodes in the tree on the basis of the groups in the required schemas if they share the same parent. However it has been stated in [9] that the split up operation need not be performed since the original group was already validated with respect to the original integrated interface. Hence this step is not considered for the development of the algorithm, in addition the split operation will result in violating the integrity of the unified interface. Figure. 1 depicts the two steps remove and collapse given an initial integrated tree.

Figure 1(a) shows an integrated tree with five leaf nodes 'a', 'b', 'c', 'd' and 'e'. Assuming that the interfaces that are required contain only the attributes 'a', 'b' and 'd', the leaf nodes 'c' and 'e' will then have to be removed. Figure 1 (b) shows integrated tree with the leaf nodes 'c' and 'e' removed, this is the removal step. Since 'e' was a child of the root it's removal causes no problem, however, removing the leaf node 'c' resulted in its sibling 'd' becoming an only child of its parent. Hence node 'd' is collapsed with its parent to become a child of the root. This is the collapse operation, depicted in Figure 1(c). On completion of the collapse operation the integrated tree obtained will have only the attributes of those interfaces that are required.

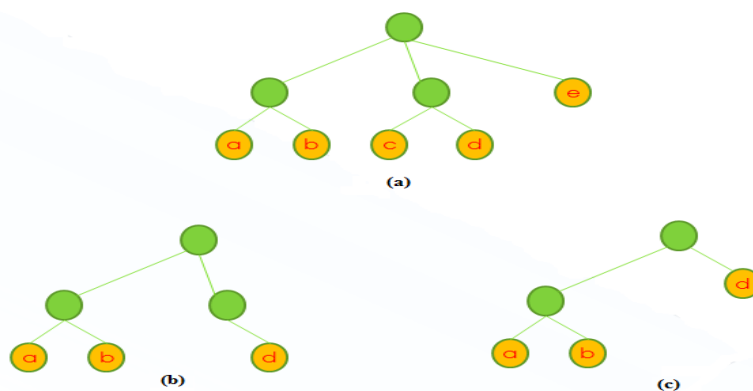


Figure 1. Pruning the Integrated Tree

**III. PRUNING ALGORITHM FOR INTEGRATED ATTRIBUTE CONSTRAINT MATRIX**

The pruning algorithm for unified constraint matrices is given in Figure 2, the unified attribute constraint matrix ‘M’ generated using the technique in [7] as well as the set ‘Q’ of all the interfaces that are required by the user to be integrated is given as the input. The output of this algorithm will be the matrix ‘QI’ containing only the attributes of those interfaces required by the user. The ordered tree for this matrix QI can be easily generated using the technique described in [7]. The algorithm is followed by an example of its operation on attribute constraint matrices based on the integrated tree in Figure1.

---

Algorithm **PruneUnifiedMatrix** (M, Q)

---

**Input:**

M: Unified attribute constraint matrix of all interfaces in domain  
 Q: Set of query interfaces specified by user.

**Output:**

QI: Integrated attribute constraint matrix of user specified matrices

---

1. QI=M
  2. Generate set SET1 of all attributes in QI.
  3. Generate set SET2 of all distinct attributes in Q.
  4. Determine groups of siblings Sib1, Sib2, Sib3.....Sibn in QI.
  5. SETa=SET1-SET2.
  6. Remove from QI all attributes in SETa.
  7. **If** any attribute or combinations of attributes ‘att’ in SETa  $\in$  {Sib1~Sibn}
    - A= {Sibi - att}
    - If** |A|=1 // lone attribute after siblings are removed
      - b=QI [A,A] // current depth of lone attribute
      - QI [A, A] =1 // set depth of lone attribute in QI to 1
    - for (i=0;i<sizeofQI;i++)
    - QI [i, A] =QI [i, A]-(b-1) // update distance of all other attributes to lone attribute
- End If**
- 

Figure 2. Pseudo code of algorithm PruneUnifiedMatrix

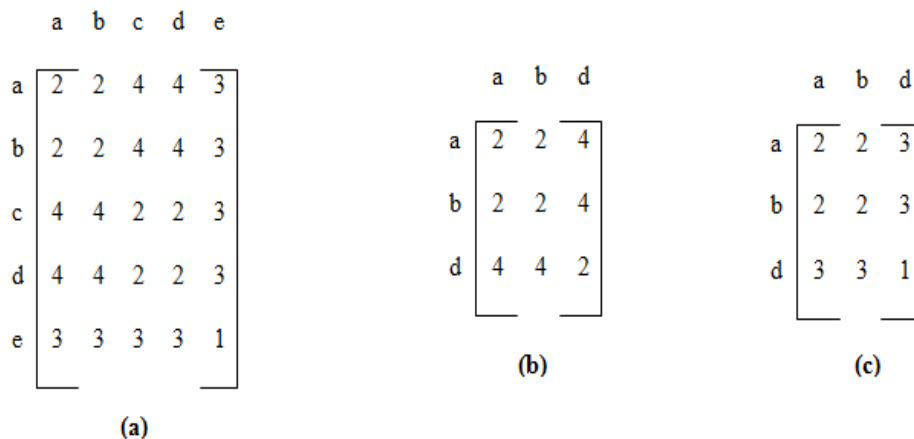


Figure 3. Attribute constraint matrix values at different stages in the algorithm PruneUnifiedMatrix



Figure 3(a) shows the attribute constraint matrix corresponding to the tree in Figure 1(a), matrix is generated using the technique described in [7] with the diagonal values representing the depth of the node from the root, and the remaining values representing the number of edges to be travelled between the corresponding nodes. Figure 3(b) shows the matrix after the remove operation in line 6 of the algorithm in Figure 2 has been executed, the rows and columns corresponding to the attributes 'c' and 'e' have been removed which are assumed to be those attributes not present in the interfaces required by the user to be integrated. Figure 3(c) shows the final attribute constraint matrix after the entire algorithm in Figure 2 has been executed. After the remove operation the attribute 'd' was found to be the lone attribute since its sibling 'c' was removed, its depth has been set to 1 and its distance to the remaining attributes has been reduced by the extent to which its depth has to reduce to become 1. Notice that the matrix in Figure 3(c) will be the same as the matrix drawn from Figure 1(c) using the technique in [7], hence showing that this algorithm works correctly on the attribute constraint matrices suggested by [7].

#### IV. RESULT ANALYSIS

The proposed algorithm in Figure 2 was implemented in the Java language with the input as the examples given in [7], the satisfaction rate of hierarchical and precedence constraints defined in [7] was satisfactory, however the satisfaction rate of the group constraints [7] was found to reduce by a small extent when compared to the group constraint satisfaction rate of the initial integrated interface, this is due to the existence of lone siblings after the removal process. Overall this algorithm for customizing the integrated query interface as per the users' preference does not violate the constraints on the structures of the constituent schemas

#### V. CONCLUSIONS

Large scale integration of query interfaces is a widely studied problem in deep web data integration. Representation of the query interfaces as attribute constraint matrices was found to be capable of capturing all the constraints imposed on a query interface. The final generated unified matrix facilitated the development of a unified interface that maintained all the characteristics of the constituent matrices. Customizing this integration according to user preference involved developing an algorithm that could be applied on this unified matrix such that it will ultimately contain only the attributes of those interfaces specified by the user. Java programming language was used to implement the core algorithm and a unified matrix was generated from the given input matrices using the proposed algorithm which did not violate the constraints imposed by the constituent schemas.

#### REFERENCES

- [1.] [1] H. He, W. Meng, C. Yu and Z. Wu, *WISE-Integrator: an automatic integrator of web search interfaces for e-commerce*, VLDB'03 (2003).
- [2.] [2] H. He, W. Meng, C. Yu and Z. Wu, *Automatic integration of web search interfaces with WISE-Integrator*, The VLDB Journal 13 (2004) 256–273.
- [3.] [3] H. He, W. Meng, C. Yu and Z. Wu, *WISE-Integrator: a system for extracting and integrating complex web search interfaces of the Deep Web*, VLDB'05 (2005) 1314–1317.
- [4.] [4] W. Wu, C. Yu and A. Doan, *Merging Interface schemas on the Deep Web via clustering aggregation*, ICDM'05 (2005).
- [5.] [5] E. Dragut, W. Wu, P. Sistla, C. Yu and W. Meng, *Merging source query interfaces on web databases*, ICDE'06 (2006) 1–10.
- [6.] [6] F. Yuan, L. Han and Y. Wei, *A Deep Web interface integration approach based on keyword matching and similarity computing*, Journal of Computational Information Systems 6 (2009) 1569–1576.
- [7.] [7] Y. Li, Y. Wang, P. Jiang and Z. Zhang, *Multi-Objective Optimization Integration of Query Interfaces for the Deep Web based on Attribute Constraints*, Data & Knowledge Engineering 86 (2013) 38-60
- [8.] [8] Sherlin et al., *A Study on the Existing Techniques for Query Interface Integration*, International Journal of Advanced Research in Computer Science and Software Engineering 3(12), December - 2013, pp. 330-333
- [9.] [9] E.C. Dragut, F. Fang, C. Yu, W. Meng, *Deriving customized integrated web query interfaces*, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Workshops, 2009, pp. 685–688.

# Analysis of Image Segmentation Methods Based on Performance Evaluation Parameters

Monika Xess<sup>1</sup>, S. Akila Agnes<sup>2</sup>

<sup>1</sup> PG Student, Department of Computer Science and Engineering, Karunya University, Tamil Nadu, India,

<sup>2</sup> Assistant Professor, Department of Computer Science and Engineering, Karunya University, Tamil Nadu, India,

## ABSTRACT

Image segmentation is an important technology for image processing which aims at partitioning the image into different homogeneous regions or clusters. Lots of general-purpose techniques and algorithms have been developed and widely applied in various application areas. However, evaluation of these segmentation algorithms has been highly subjective and a difficult task to judge its performance based on intuition. In this paper image segmentation using FCM, Region Growing and Watershed algorithms is performed and segmentation results of these techniques are analyzed based on four performance metrics GCE, PSNR, RI and VoI. This analysis provides an overview that on what parameters different image segmentation techniques can be evaluated at best.

**Keywords:** FCM, Region Growing, Watershed, GCE, PSNR, RI, VoI.

## I. INTRODUCTION

Image segmentation is a technique of partitioning an image into several homogeneous segments or clusters based on measurements taken from the image. The output of image segmentation is a set of regions that combine to form the entire image. Some of the practical applications of image segmentation are: Content-based image retrieval, Machine vision, Medical imaging, Recognition tasks, traffic control systems, video surveillance. Several general-purpose algorithms and techniques have been developed for image segmentation. There are various methods of image segmentation such as clustering based (FCM, K-means) methods, region based methods (region growing, region splitting, region merging), watershed, edge detection method, neural networks and thresholding.

In this work segmentation algorithms chosen for analysis are FCM (Fuzzy C-Means), Region Growing and Watershed. FCM is an unsupervised segmentation algorithm that is based on the idea of finding cluster centers by iteratively adjusting their position and evaluation of an objective function. The iterative optimization of the FCM algorithm is essentially a local searching method, which is used to minimize the distance among the image pixels in corresponding clusters and maximize the distance between cluster centers [1]. FCM algorithm has long been a popular image segmentation algorithm. Region growing is a simple region-based image segmentation method. In [2] region growing is a procedure that groups pixels or sub regions into larger regions based on predefined criteria for growth. The basic approach is to start with a set of "seed" points and from these grow regions by appending to each seed those neighboring pixels that have predefined properties similar to the seed.

Watershed transform is a powerful tool that is based on the object's boundary and finds local changes for image segmentation. According to [3] watershed segmentation method is based on watershed transform. This method aims to find catchment basins, which define border between two objects. If water falls into these basins, level of the water rises until neighbor basins share the same level. So output of the algorithm is a hierarchy of catchment basins. The key point is to find most discriminative basins, since most discriminative basins are the basins that separate two different objects. The performance evaluation parameters used here for evaluation of image segmentation quality are namely PSNR (Peak signal to noise ratio), RI (Rand Index), VoI (Variation of Information) and GCE (Global Consistency Error).

## II. DESCRIPTION OF DIFFERENT IMAGE SEGMENTATION TECHNIQUES

### 2.1 Fuzzy C-Means (FCM) Algorithm

**Input:** image, cluster number.

**Output:** segmented image, cluster center and objective function [1].

1. Initialize the fuzzy cluster number, F and the cluster centers,  $c = \{c_1, \dots, c_j, \dots, c_f\}$   
Set iteration time  $q=0$ ;
2. while  $\|W^{(q)} - W^{(q-1)}\| \geq \epsilon$  do
3.     for  $j = 1$  to  $F^{th}$  cluster do
4.         for  $I=1$  to  $N^{th}$  image pixel do
5.             Calculate  $u_{ij}$ , i.e. the membership of pixel  $x_i$  to the  $j^{th}$  cluster;
6.             if  $\|x_i - c_j\| = 0$  then
7.                  $u_{ij} = 1$ ;
8.                 reset other membership of pixel i to 0;
9.             end if
10.         end for
11.     Update  $c_j$  (cluster center);
12.     end for
13.     Calculate objective function  $W^{(q)}$ ;
14.      $q=q+1$ ;
15.     end while

The objective function is given as:

$$W^{(q)} = \sum_{j=1}^F \sum_{i=1}^N u_{ij}^m \|x_i - c_j\|^2 \quad (2.1)$$

where N is the number of image pixels,  $u_{ij}$  is the membership of pixel  $x_i$  to the  $j$ th cluster,  $m$  is a constant that defines the fuzziness of the membership,  $\|x_i - c_j\|$  is the Euclidean distance between  $x_i$  and  $c_j$  [1].

The membership function is defined as :

$$u_{ij} = \frac{\|x_i - c_j\|^{-2/m-1}}{\sum_{k=1}^F \|x_i - c_k\|^{-2/m-1}} \quad (2.2)$$

The value of  $m$  is manually set and mostly  $m=2.0$  is used.

The cluster center in FCM algorithm is defined as:

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (2.3)$$

### 2.2 Region growing algorithm

Basic function of region growing algorithm is to segment the input image using  $n$  number of seed points. The basic steps followed in seeded based region growing technique includes two main points, firstly selection of initial seed point and secondly growing formula based on stopping criterion [5].

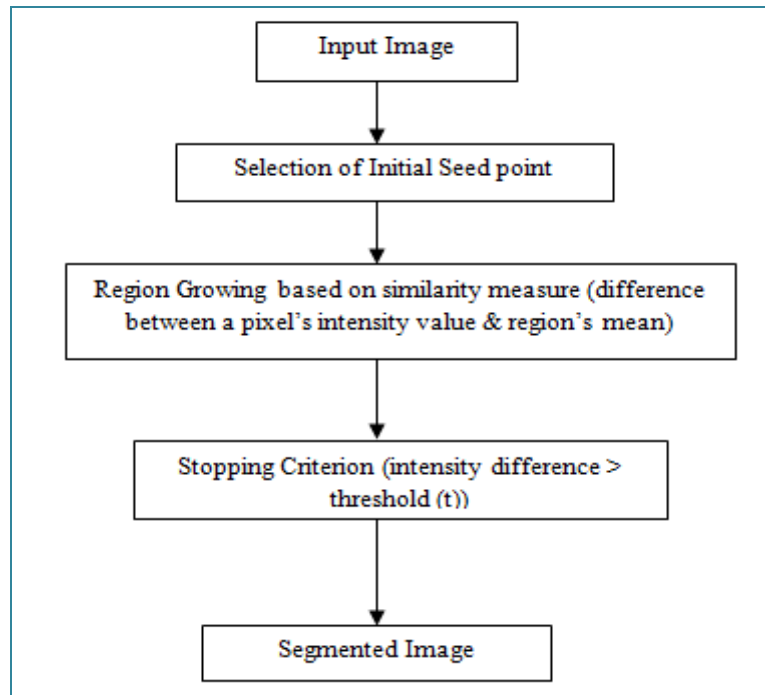


Figure1. Flow chart of Region Growing algorithm.

### 2.3 Watershed Algorithm

In this paper Marker-Controlled Watershed Segmentation algorithm is presented. The steps involved in this algorithm are: Firstly we compute a segmentation function, then compute foreground markers, after that we compute background markers and then modify the segmentation function so it only has minima at the foreground and background marker locations, finally compute the watershed transform of the modified segmentation function [4].

### 2.4 GCE (Global Consistency Error)

The Global Consistency Error (GCE) measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related are considered to be consistent, since they could represent the same image segmented at different scales. The formula for GCE is as follows,

$$GCE = \frac{1}{n} \min \{ \sum_i E(s1, s2, pi), \sum_i E(s2, s1, pi) \} \quad (2.4)$$

where, segmentation error measure takes two segmentations S1 and S2 as input, and produces a real valued output in the range [0:1] where zero signifies no error. For a given pixel pi consider the segments in S1 and S2 that contain that pixel [6].

### 2.5 PSNR (Peak Signal to Noise Ratio)

PSNR represents region homogeneity of the final partitioning. The higher the value of PSNR the better is segmentation. PSNR in decibels (dB) is computed by using [7]:

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MAE} \right) \quad (2.5)$$

MAE is the mean absolute error of the segmented image computed as follows

$$MAE = \frac{1}{MN} \sum \sum |F(i, j) - f(i, j)| \quad (2.6)$$

F (i, j) - segmented image, f (i, j) - source image that contains M by N pixels.

### 2.6 RI (Rand Index)

The Rand index (RI) between test and ground-truth segmentations S and G is given by the sum of the number of pairs of pixels that have the same label in S and G and those that have different labels in both segmentations, divided by the total number of pairs of pixels [8]. In [9] the RI is given as:

$$RI = \frac{a+b}{a+b+c+d} = \frac{\binom{n}{2} [0.5\{\sum_i (\sum_j n_{ij})^2 + \sum_j (\sum_i n_{ij})^2\} - \sum \sum n_{ij}^2]}{\binom{n}{2}} \quad (2.7)$$

where, a, the number of pairs of elements in S that are in the same set in U and in the same set in V ; b, the number of pairs of elements in S that are in different sets in U and in different sets in V; c, the number of pairs of elements in S that are in the same set in U and in different sets in V; d, the number of pairs of elements in S that are in different sets in U and in the same set in V.  $n_{ij}$  is the number of objects in the  $i$ th cluster in U and  $j$ th cluster in V, and  $\binom{n}{2}$  is the binomial coefficient, which gives the number of distinct pairs found in a set of  $n$  objects.

**2.7 VoI (Variation of Information)**

The Variation of Information (VoI) metric defines the distance between two segmentations as the average conditional entropy of a segmentation given the other, and thus roughly measures the amount of randomness in a segmentation which cannot be explained by the other [5]. Lower the VoI value better is the result. In [10] VoI is defined as:

$$VI(c, c') = H(c) + H(c') - 2I(c, c') \quad (2.8)$$

where,  $H(c)$  and  $H(c')$  are the entropies associated with cluster  $c$  and  $c'$ ;  $I(c, c')$  is the mutual information between the associated random variables.

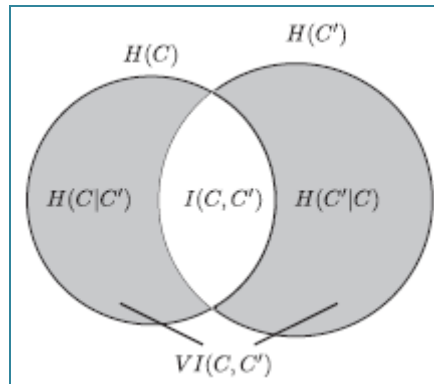


Figure2. The variation of information and related quantities

**III. IMPLEMENTATION DETAILS AND RESULT ANALYSIS**

The three image segmentation algorithms discussed in this paper are implemented in MATLAB (R2010a) and their results are evaluated using four quantitative performance metrics namely, GCE, PSNR, RI and VoI.

For better segmentation results the values of PSNR (Peak Signal to Noise Ratio) and RI (Rand Index) should be higher and on the other hand the values of GCE (Global Consistency Error) and VoI (Variation of Information) should be low. The image segmentation algorithms in this paper are applied to natural images from BSD database, where test and ground-truth images are used to calculate GCE, RI and VoI values [11]. The results of segmentations performed by three different algorithms are described below taking three natural images flower, airplane and kid. Four performance metrics are then calculated separately for each algorithm and performance analysis is done.

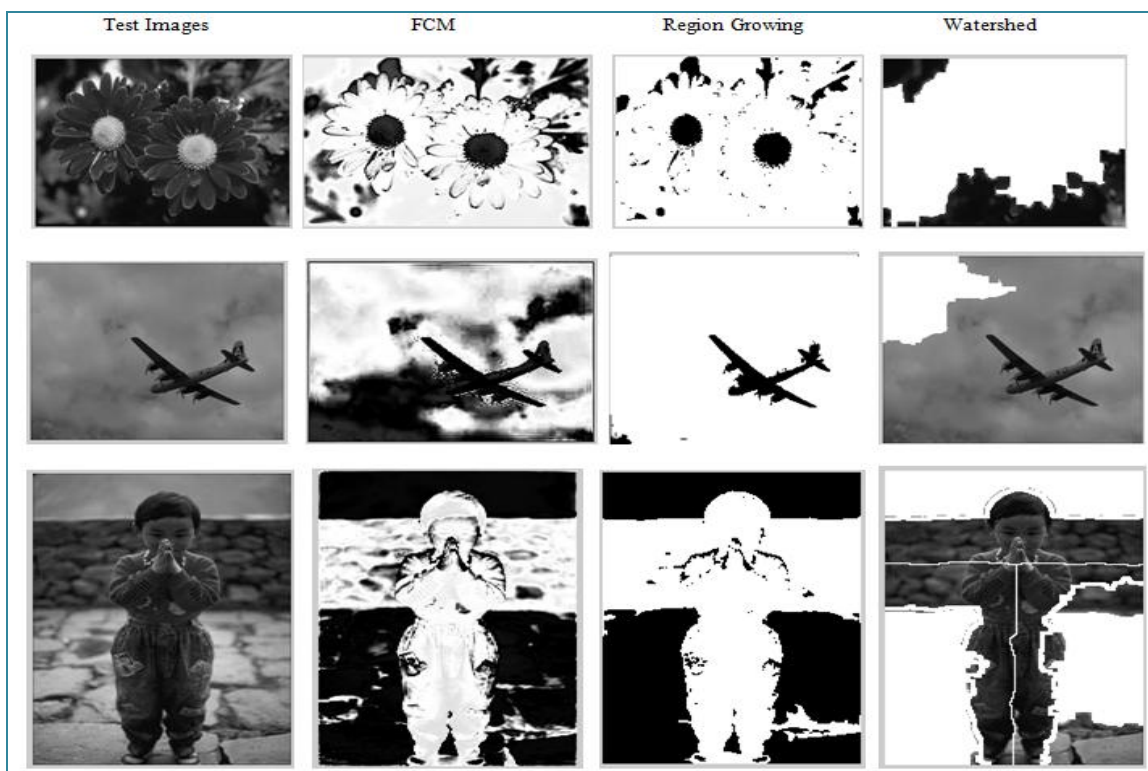


Figure 3. Segmentation results of FCM, Region Growing and Watershed algorithms performed on three natural images Flower, Airplane and Kid.

Table1. PSNR, RI, VoI and GCE values calculated for FCM algorithm

Test Image	PSNR	RI	VoI	GCE
<b>Flower</b>	17.1785	0.9367	0.2079	0
<b>Airplane</b>	12.6525	0.9566	0.1536	0
<b>Kid</b>	13.8960	0.9496	0.1732	0

Table 2. PSNR, RI, VoI and GCE values calculated for Region Growing algorithm

Test Image	PSNR	RI	VoI	GCE
<b>Flower</b>	56.0955	0.7690	0.6987	0.0633
<b>Airplane</b>	60.0042	0.8710	0.4581	0.0429
<b>Kid</b>	57.2841	0.5001	1.1724	0.0504

Table 3. PSNR, RI, VoI and GCE values calculated for Watershed algorithm

Test Image	PSNR	RI	VoI	GCE
<b>Flower</b>	56.2281	0.9367	0.2079	0
<b>Airplane</b>	59.3651	0.9566	0.1536	0
<b>Kid</b>	46.9179	0.3208	2.1163	0.0501

Table 4. Comparative analysis of FCM, Region growing and Watershed algorithm on three images (average value of parameters), bold values represent highest, italic second highest value.

	FCM	Region Growing	Watershed
<b>PSNR</b>	14.5757	<b>57.7946</b>	<i>54.1704</i>
<b>RI</b>	<b>0.9477</b>	<i>0.7134</i>	0.7380
<b>VoI</b>	<b>0.1782</b>	<i>0.7764</i>	0.8259
<b>GCE</b>	<b>0</b>	0.0522	<i>0.0167</i>

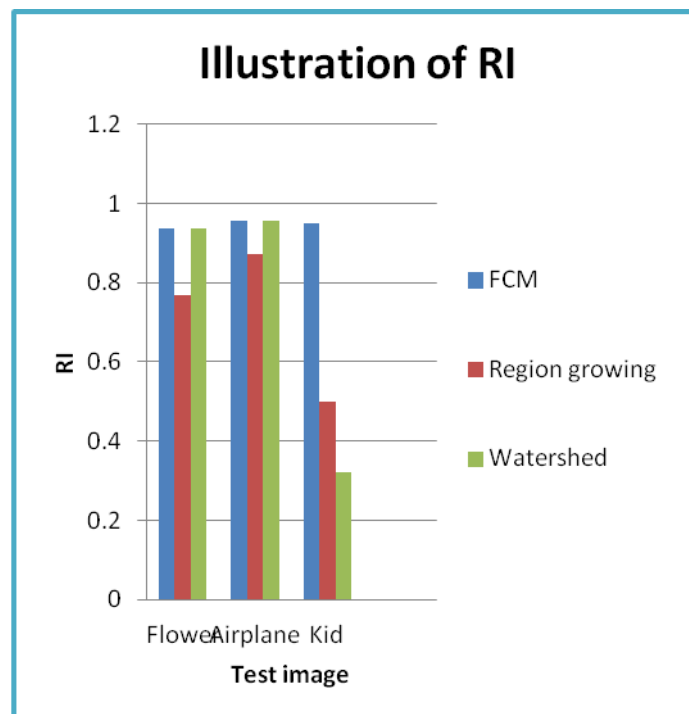
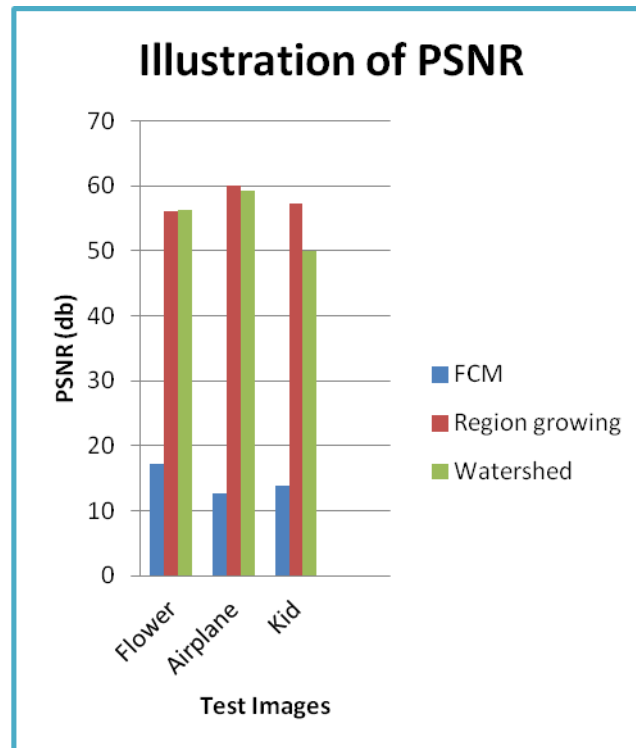


Figure4. Illustration of PSNR and RI values for FCM, Region growing and Watershed algorithm

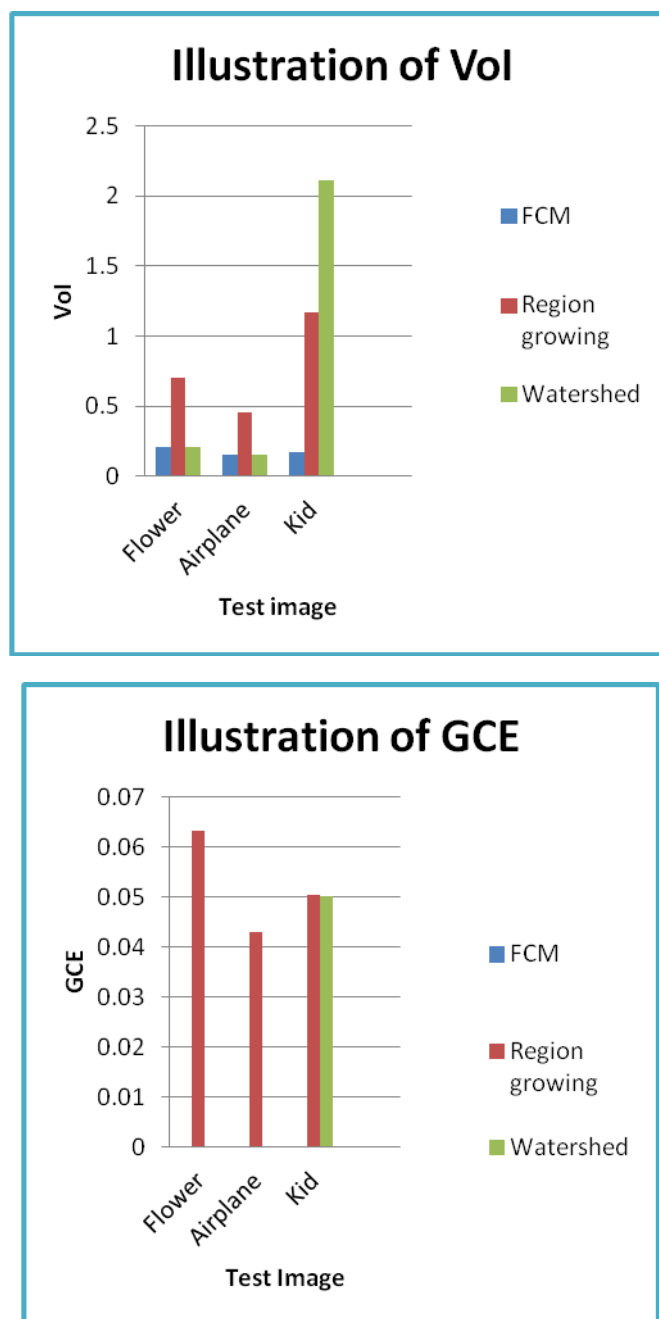


Figure5. Illustration of VoI and GCE values for FCM, Region growing and Watershed algorithm

#### IV. CONCLUSION

Three image segmentation algorithms namely FCM, region growing and watershed are discussed and their segmentation results are evaluated using four quantitative measures. Comparative analysis shows that region growing algorithm has more de-noising capability with highest PSNR value. Measures suitable for evaluating FCM algorithm are Rand Index (RI), VoI and GCE, where GCE (Global Consistency Error) for FCM is zero indicating better segmentation result. Also PSNR is suitable for evaluating Watershed algorithm. Thus, different image segmentation algorithm can be evaluated with suitable performance metric based on its application.



## REFERENCES

- [1] Tan .S. K., Lim .H.W, Isa .M.N (2013), "Novel initialization scheme for Fuzzy C-Means algorithm on color image segmentation", *Applied Soft Computing* 13, 1832-1852.
- [2] R.C. Gonzalez, R.E. Woods (2011), "Digital Image Processing", Third Edition, ISBN 978-81-317-2695-2.
- [3] T. Berber, A. Alpkocak, P. Balci , O. Dicle (2013), "Breast mass contour segmentation algorithm in digital mammograms", *Computer methods and programs in Biomedicine* 110, pages 150–159.
- [4] M. A. Hamdi (2011), "Modified Algorithm marker-controller watershed transform for Image segmentation Based on Curvelet Threshold", *Canadian Journal on Image Processing and Computer Vision* Vol.2 No.8.
- [5] Jianping et al. (2005), "Seeded region growing: an extensive and comparative study", *Pattern Recognition* 26, 1139-1156.
- [6] B. Sathya and R. Manavalan (2011), "Image Segmentation by Clustering Methods: Performance Analysis", *International Journal of Computer Application*, 0975-8887.
- [7] Sowmya. B. and Rani Sheela. B. (2011), "Colour image segmentation using fuzzy clustering techniques and competitive neural network", *Applied Soft Computing* 11, pages 3170–3178.
- [8] Arbela, X.P. ez, M. Maire, C. Fowlkes, J. Malik (2011), Contour detection and hierarchical image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, pages 898–916.
- [9] W. Liu, B. Wang, J. Glassey, E. Martin, J. Zhao (2009), "A novel methodology for finding the regulation on gene expression data", *Progress in Natural Science* 19, pages 267–272.
- [10] M. Meila (2007), "Comparing clusterings—an information based distance", *Journal of Multivariate Analysis* 98, 873 – 895.
- [11] D. Martin, C. Fowlkes, D. Tal and J. Malik (2001), "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics", *Proc. 8th Int'l Conf. Computer Vision*, Vol. 2, pages 416-423.

# Controlling Computer Operations using Brain-Wave Computing

Shanmugapriya.B<sup>1</sup>, Akshaya.T<sup>2</sup>, Kalaivani.K<sup>3</sup>, Anbarasu.V<sup>4</sup>

Department of Information Technology, Jeppiaar Engineering College, Chennai, India

## Abstract:

People interact with computer using devices that have been created to serve a specific purpose. For generations, humans are fascinated about the idea of communicating with machines through devices that can peer into person's mind. Such idea motivated to the recent advancements in the field of artificial intelligence and cognitive neuroscience to provide the ability to interact with machine through brain. The proposed work is an Electroencephalography (EEG) based biomedical signal processing system to perform computer operations by manipulating the brain activity. The users have to explicitly manipulate the brain activity to produce signals that can be used to operate the computer. The brain waves are obtained with the help of scalp electrodes. EEG signals collected are then processed to interpret the command and execute the desired task. The real time implementation requires training the computer according to one's thoughts and actions through neural networks.

**Keywords:** Artificial Intelligence, Cognitive Neuroscience, Electroencephalography, Biomedical signal processing, Scalp electrodes, Neural Networks.

## I. INTRODUCTION

Human-Computer Interface (HCI) allow human to work in conjunction with machine. Some of the HCI devices like EON mirror and Kinect work on the basis of gesture recognition and the most popularly used touch screen devices require tactile input. In the proposed work, brain thoughts are monitored to handle a machine making it narrower and specific to Brain-Computer Interface (BCI) from the broader Human-Computer Interaction (HCI). The researchers in BCI recognize the need for system, that makes BCI more user-friendly, real-time, manageable and more suitable for all people, expanding its region of use with clinical patients and those who are physically challenged.

The Electroencephalography (EEG) signals are generated in the brain through the voltage difference of ions moving through the neurons. EEG signals are nothing but the brain activity in the form of electro-voltaic waves. Originally, the signals were used for monitoring the brain activity of the patients. Later on, it advanced to place its footprints in BCI.

Several systems use EEG signals to identify the person, provides an aid to disabled people and there are a quantity of systems which beat the idea of the previous two, allowing gamers to use it beneficially to bag high scores according to their focusing level. Very little work has been done in this area, focusing more on medical purposes, not on overall computer operation. People wonder why to constrict the use of BCI only to particular purpose and why not for all the operations that can be done with a computer. The proposal is to insist on the use of brain thoughts as input to the Computer to operate it in a more efficient and easier manner.

The proposed system provides an alternative to other input devices, because the brain thoughts of the user are taken as input to the computer. The inputs are in form of EEG signals which are processed and then introduced to artificial neural network (ANN). ANN is a machine learning approach which replicates human brain and contains a number of artificial neurons. It increases the privacy and security of the user thus making the system more responsible to the authenticated person and resistive to the impostor.

## II. EXISTING SYSTEMS

EEG signals are recommended to use by the BCI researchers to lift the ease of operating a system in various regards. Kusuma Mohanachandra and the group have used EEG signals in Brain Computer Interface as a new modality for Person Authentication and develop a screen lock application [1] that will lock and unlock the computer screen at the users will. The brain waves of the person, recorded in real time are used as password to unlock the screen. Use of EEG signals in BCI makes it more secure as nobody knows what the user is thinking of. It provides more privacy that convinces you to use it for authentication. Even though there are already graphical and click - based passwords, use of EEG in authenticating a person is a step ahead in providing security. Sébastien Marcel created a statistical framework that has also been created based on Gaussian Mixture Models and Maximum A Posteriori model adaptation [2], successfully applied to speaker and face authentication. Using EEG signals eliminates the possibilities of occurrence of shoulder surfing and other password cracking attacks. It allows the user to be trained to work with the authentication system and then be tested for the identity. The system they implemented also improves incremental learning. But the problem is hours of training are needed for both authentications.

Katherine developed a system to identify people using EEG signals generated during their imagined speech [3]. With the help of the model, subjects are identifiable to 99.76% accuracy. But, there are chances that people can't imagine their speech with the same characteristics every time. Cheng He suggested an idea of user authentication with more accuracy and robustness [4] using EEG signals. He addressed two major problems associated with EEG biometrics. One is that the large EEG features size and the relatively limited EEG data size, make it difficult to train a robust model; the other is that the signals from EEG scalp may not be reliable in many situations.

As an attempt to make physically challenged people more independent, Jonathan and his crew used EEG signals to aid them to control their movement. A restoring function is provided to those with motor impairments providing the brain with a new, non-muscular communication and control channel [5], a direct brain-computer interface (BCI) for conveying messages and commands to the external world. Joseph developed a mobile imaging approach [6] that provides a tight coupling between human physical structure with cognitive processing and the role of supraspinal activity during control of human stance and locomotion.

Apart from the above works, researcher I-Jan Wang developed a wearable, wireless EEG- based headset [7] to win the shot in gaming, monitoring the focusing level of the user using the EEG signals. Chin-Teng Lin eliminated the use of conduction gel required by the use of foam based dry electrodes [8] instead of ordinary wet- scalp sensors that is used normally, to reduce the skin-electrode contact impedance. Qiang Wang has proposed Neuro-feedback games [9] and design algorithms to implement the 2D and 3D concentration games. Neuro-feedback games are assessed by the healing effect of ADHD (Attention Deficit Hyperactivity Disorder) with significant EEG distortion.

From a thorough investigation on BCI researches, the following limitations are spotted out in the current BCI systems. The BCI systems which are in use today are developed in such a way that they can serve only a specific purpose. Because of the fact that only an intended purpose can be served with the help of existing system, the role of devices that are used for interacting with computer cannot be avoided completely. EEG headset is needed in addition to already connected traditional interacting devices. The user may find it difficult to deal with multiple interacting devices simultaneously. There are possibilities that the user may get bored by performing the same kind of action all the time.

### **III. PROPOSED SYSTEM**

The system proposed will create a new generation in BCI which eliminates the need for any assisting devices to operate a computer. Using this system, computer can be operated with the help of our thoughts. All the system needs as input is the user's brain waves which are the results of some brain activity i.e. thoughts. The brain waves of the user are collected in the form of Electroencephalography (EEG) signals that can be processed further in MATLAB. The EEG signal of the user is used as input to computer; it lifts the privacy of the user to a superior level.

The pattern of collected signal is to be analyzed and the features are to be extracted. Before feature extraction, the collected signal must be amplified so that it would be easier to extract features without any error or loss of information. After the signal strength is boosted up, the artifacts in the signal must be detected and removed.

The refined EEG signal is then processed to extract features which help to identify the command of the user. For the identification of the command, one must classify the signal to its corresponding output. The most important step is to create a model of the neural system of the user's brain. The neural network must undergo training in order to understand the user's thoughts and other brain activities. With the help of back propagation algorithm [11] the training of neural networks is done. Once the command is interpreted from the signal, the logic corresponding to implement the command is executed to perform the intended task. The real-time implementation is entirely based on how the training is done and how quick the system can identify the command that is needed to be executed.

#### **1.1. Advantages of Proposed System**

Many physically challenged people have a necessity to use computers every day. It is obvious that these people will face difficulty in interacting with the computer. The system proposed helps the physically challenged people to work with computer in a much easier way. The system has been designed to reduce the fatigue of normal computer users while having to deal with various external devices attached to the computer, in order to operate it. The idea used in the proposed system enhances the resistivity of the system to inexperienced users and reduces the operational difficulty faced by the computer users. Hereby, with the use of proposed system, we can provide an alternative approach to handle a computer in the absence of traditional devices like mouse, keyboard.

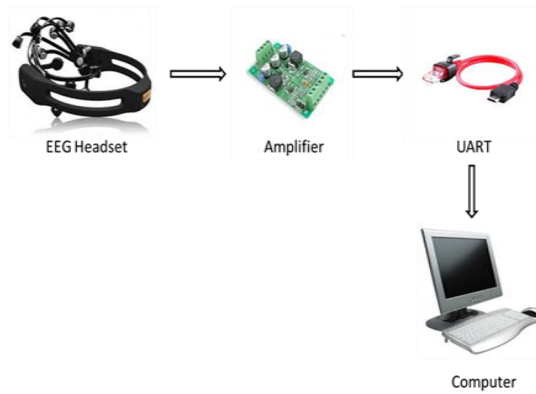


Figure 1. General Architecture

#### IV. SYSTEM IMPLEMENTATION

The proposed system consists of four different parts that are needed to be implemented. Figure 2 shows what are process are needed to be implemented for our system to work.

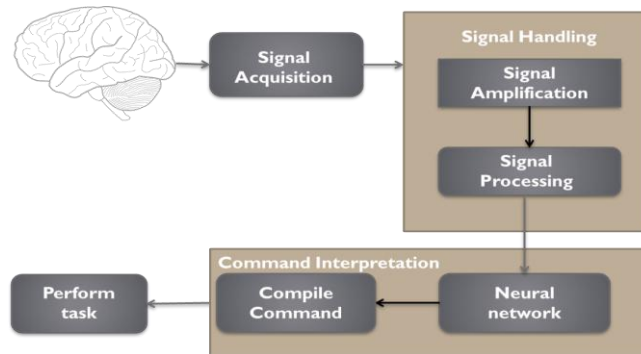


Figure 2. System Architecture

##### 1.2. Signal Acquisition

Signal acquisition involves collecting the electroencephalography signals generated in the human brain. EEG signals are generated due to the electrical activity on the brain scalp. Under the human scalp, there is a flow of ions through the neuron network that occurs due to user’s thoughts and actions. The voltage fluctuations due to the ionic current in the neural network of human brain are recorded as EEG signal. EEG signals can be recorded in two ways: one is through surgical procedure and other is done by scalp electrodes. The most preferred is the second method in which an EEG headset is used to gather brain waves of the user. EEG signals are nothing but voltage fluctuations that occur due to brain activity.

The user has to place the headset on the scalps in order to record the EEG signals. EEG headsets are designed by integrating 16 highly sensitive sensors in such a way that it senses the voltage fluctuations in the human scalp. The electrodes in the headset are to be placed in the standard positions [10] as prescribed by International 10-20 system. International 10-20 system is the internationally accepted idea of location specification of the electrodes to obtain the EEG signals generated in the human brain. The “10” and “20” refer to the fact that the actual distance between the electrodes placed on the scalp are 10% or 20% of the total front-back or left-right distance of the skull. Before placing the sensors on the scalp, it is advised to apply conductive gel in order to reduce the electrode-scalp impedance.

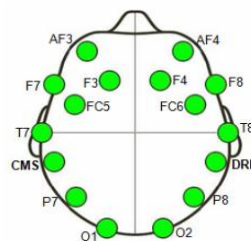


Figure 3. Standard location of sensors as per International 10-20 system

EEG signals have Gamma, Alpha, Beta, Theta and Delta wave patterns. Alpha waves are generated when the user is in a state of physical and mental relaxation and completely aware of what is happening around. Beta waves are emitted when the person feels agitated, stressed or afraid whereas theta waves get emitted during reduced consciousness of the subject. And delta waves are generated during deep sleep or unconsciousness. Alpha waves are used to implement the proposed system. During the recording of the signals, sometimes the amplitude of the signal reaches an observable length. The observable peak is called as Event Related Potential (ERP). The ERP defines that an event has occurred in brain or a thought has been generated in the brain. The most used is the P300, where the EEG signal peak reaches after 300 milliseconds after the event has occurred.

### 1.3. Signal Handling

The signal acquired from brain is very weak (about 30 – 100  $\mu\text{V}$ ). Hence it is necessary to amplify this signal before processing. Signals from the EEG headset are amplified using EEG amplifier kit. Amplifying is just boosting up the strength of the signal. Amplifiers are used to amplify the signals i.e. to improve the strength of the signal and can reduce the noise present in the signal. EEG 100C amplifier is an Electroencephalogram amplifier that amplifies bio-electric potentials associated with neural activity of the brain. The output delivered by the amplifier can be switched between normal EEG output and Alpha wave detection. The amplifier used to amplify EEG signals is voltage amplifier since these signals are the measured in voltage potentials. Voltage amplifiers are commonly available type of amplifier which amplifies input voltage to a larger output voltage. The input impedance of amplifier is high whereas the output impedance is low. The amplified signal must be given to the computer through an UART (Universal Asynchronous Receiver/Transmitter).

The EEG data is usually mixed with huge amounts of useless data produced by physiological artifacts that masks the EEG signals. Artifacts are signals that are recorded by an EEG headset, but are not of cerebral origin i.e. those are not generated due to brain activity. In most cases, the amplitude of artifacts is more than the amplitude of EEG signals. Hence, there are chances that artifacts can suppress significant features from EEG. Artifacts may be physiological (generated from the person) or non-physiological (generated from the environment or equipment). Common artifacts may include power-line artifact, eye blinking, eye movement, respiratory artifact, electrode popping, sweat artifact or due to breakage of electrode. AAR (Automatic Artifact Removal) [12] automatically removes artifacts from EEG data based on blind source separation and other various algorithms. The AAR toolbox is implemented as an EEGLAB plug-in in MATLAB and was used to process our EEG data subset on two stages: Electrooculography (EOG) removal using the Blind Source Separation (BSS) algorithm then Electromyography (EMG) removal using the same algorithm.

After artifact detection and removal, the signal is processed by Fast Fourier Transform (FFT). FFT is used to eliminate all the unnecessary altering signals in the brain wave i.e. EEG signal. The FFT is applied to the signal and its amplitude is used for further process.

*DFT(FFT):*

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j\left(\frac{2\pi}{N}\right)nk} \quad (k = 0, 1, \dots, N-1)$$

*IDFT(IFFT):*

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cdot e^{j\left(\frac{2\pi}{N}\right)nk} \quad (n = 0, 1, \dots, N-1)$$

**Figure 4. FFT and its inverse**

where  $x(n)$  is the input signal values and  $e$  is an exponential constant. The above method helps to retain the range of the signal, now we need to normalize it. A simple sampling theorem has been used to normalize the signal. Based on the information needed to be retained and those afforded to be lost, sampling frequency must be chosen.

$$f_s \geq 2 f_m$$

where  $f_s$  – Sampling frequency and  $f_m$  – frequency of the modulating signal. It is recommended to set the sampling frequency equal to Nyquist rate. Nyquist rate is the minimum frequency at which the signal is to be sampled to avoid any loss of information.

$$f_s = 2 f_m$$

After sampling, the samples are to be quantized i.e. the samples are rounded-off to the nearest amplitude value.

#### 1.4. Command Interpretation

Interpretation of commands in the system is done by the neural network, which is a type of signal classifier. The neural network is specified by the neuron model, architecture, and learning algorithm. The network must be trained and should be able to behave correctly on new instances of learning tasks. A neuron is a basic information processing unit. The architecture is linked with the learning algorithm used for training the system.

Multilayer perceptron is the type of neural network consists of an input layer, possibly a number of hidden layer, and one output layer. The output of each node in one layer connects to the input of the next layer, but not within the same layer.

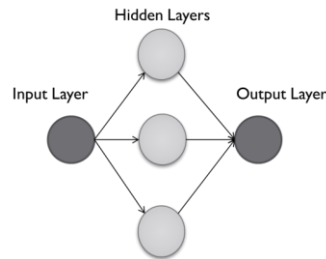


Figure 5. Neural network

Neurons are also referred to as nodes, and the lines between the nodes are called synapses. The inspiration comes from the human brain but it is a simplified version. Each neuron consists of weights, adder functions that adds all the weights and activation function which is used to limiting amplitude of the output. The boundless strength with neural networks is that they are flexible, if given the right setup.

Neural networks needs to be trained before it can be used. In case of supervised learning, training sets with sample input and the resultant output are given to the network, and then an algorithm alter the weights of the synapses so that it can map valid input to correct output. A very common algorithm for this is the back-propagation algorithm.

##### 1.4.1. Back- propagation Algorithm

The training is done using back- propagation algorithm. The algorithm searches for the weight and this can be found with the help of error value during training set. The weight is defined as the strength of connection between two nodes. If the error value is minimum compared to the previous values, then that corresponding weight value is found. Training sets are given to the algorithm which defines what user want the network to do and changes on network's weights so that, when training is finished, it will give you the required output for a particular input. Back propagation networks are ideal for simple Pattern Recognition and Mapping Tasks.

The training begins with the network, first initialised by setting up all its weights to be small random numbers say between  $-1$  and  $+1$ . Next, the input pattern is applied and the output calculated (this is called the forward pass). The calculation gives an output which is completely different to what user wants (Target), since all the weights are random. Then calculate the Error of each neuron, which is essentially: Target – Actual Output (i.e. What user wants – What user actually got). The error value is then used mathematically to change the weights in such a way that the error gets smaller. In other words, the Output of each neuron will get closer to its Target (this part is called the reverse pass). The process is repeated again and again until the error is minimal. The pseudo code for the algorithm is given below:

```

Initialize the weights;
while stopping criterion has reached do
for all example e training set do
    O = actual, output (network, e); //propagate forward
    T = wanted output for e
    Calculate error (T - O) at each neuron in the output layer
    Compute Mean Squared Error value; //propagate backward
    Compute updateweight for all weights
    Update all the weights in the network such that
        Sum-squared value of error is minimized
end for
end while

```

The Mean Squared Error (MSE) value is calculated and forms the performance index. This value reflects the effectiveness of the training done so far. The stopping criterion could either be when the MSE has reached an acceptable limit, or when the number of training cycles is attained. Once the correct signal is identified, the logic representing the particular command for that signal has to be identified. The command intended to perform is opening a notepad. During training, first as a baseline the user is asked to be in a meditated and relaxed state. For execution of the command, the user is asked to solve a math question and accordingly the training is done and that action is entitled to that operation to open the notepad. Now to test whether the system works, the user is made to solve a math question. Then after recording and processing of the EEG signal, enters to the neural network where the command is identified. The logic that perfectly matches with the command has to be compiled and the intended task will be performed.

## **V. Conclusion**

The evolution of BCI has brought up a revolution in the current technical era. The system proposed has a necessity to provide initial training to the system. Because, the system have to identify which action equals to the desired command. Commands are completely action specific i.e. the imagination of a particular action by the user. In future, the system must be modified in such a way that it can respond to the user commands automatically eliminating the need to train the systems regarding the command specification. People may have like thoughts but not completely same thoughts. Hence, using thoughts of a person for intermingling with a computer can augment the security. Also, the uniqueness of thoughts among a vast population of people would boost up privacy of the user. The biggest challenge in the proposed work is not just the implementation of the system but the training of the neural network. The system is perfect for handicapped people who find difficulties in using the system. Brain- wave computing will indeed be a new alternative for manual work needed to control the system.

## **REFERENCES**

- [1] Kusuma Mohanchandra, Lingaraju G M, Prashanth Kambli & Vinay Krishnamurthy, "Using Brain Waves as New Biometric Feature for Authenticating a Computer User in Real-time", International Journal of Biometrics and Bioinformatics (IJBB), Volume (7) : Issue (1) : 2013.
- [2] S'ébastien Marcel and Jos'é del R. Mill'an , "Person Authentication using Brainwaves (EEG) and maximum A Posteriori Adaptation", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, SPECIAL ISSUE ON BIOMETRICS, April 2006.
- [3] Katharine Brigham and B. V. K. Vijaya Kumar, "Subject Identification from Electroencephalogram (EEG) Signals During Imagined Speech", Biometrics Compendium, IEEE 2010.
- [4] Chen He, "Person Authentication Using EEG Brainwave Signals", September 2009.
- [5] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, Theresa M. Vaughan, "Brain-Computer Interface for Communication and Control", Clinical Neurophysiology, March 2002.
- [6] Klaus Gramann, Joseph T. Gwin , Daniel P. Ferris , Kelvin Oie , Tzyy-Ping Jung, Chin-Teng Lin , Lun-De Liao and Scott Makeig, "Cognition in action: Imaging body/brain dynamics in mobile humans".
- [7] Lun-De Liao, Chi-Yu Chen, I-Jan Wang, Sheng-Fu Chen, Shih-Yu Li, Bo-Wei Chen, Jyh-Yeong Chang and Chin-Teng Lin, "Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors", Journal of Neuro-Engineering and Rehabilitation 2012.
- [8] Chin-Teng Lin, Lun-De Liao, Yu-Hang Liu, I-Jan Wang, Bor-Shyh Lin, and Jyh-Yeong Chang, "Novel Dry Polymer Foam Electrodes for Long-term EEG Measurement", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 58, NO. 5, MAY 2011.
- [9] Qiang Wang, Olga Sourina, and Minh Khoa Nguyen, "EEG-based "Serious" Games for Medical applications".
- [10] Iván Manuel Benito Núñez, "EEG Artifact Detection", June 2010.
- [11] Erik Andreas Larsen, "Classification of EEG Signals in a Brain- Computer Interface System", June 2011.
- [12] Mohammad H. Alomari, Aya Samaha, and Khaled AlKamha, "Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning", International Journal of Advanced Computer Science and Applications, Vol. 4, No. 6, 2013.

# A Novel Approach for the Effective Detection of Duplicates in XML Data

Anju Ann Abraham<sup>1</sup>, S. Deepa Kanmani<sup>2</sup>

*1 PG Student, Department of Computer Science and Engineering, Karunya University.*

*2 Assistant Professor, Department of Computer Science and Engineering, Karunya University.*

## **Abstract:**

*eXtensible Markup Language is widely used for data exchange between networks and it is also used for publishing data on web. Identifying and eliminating the duplicates has become one of the challenging tasks in the area of Customer Relationship Management and catalogue integration. In this paper a hybrid technique is used for detecting duplicates in hierarchically structured XML data. Most aggressive machine learning techniques is used to derive the conditional probabilities for all new structure entered. A method known as binning technique is used to convert the outputs of support vector machine classifiers into accurate posterior probabilities. To improve the rate of duplicate detection network pruning is also employed. Through experimental analysis it is shown that the proposed work yields a high rate of duplicates thereby achieving an improvement in the value of precision. This method outperforms other duplicate detection solution in terms of effectiveness.*

**Keywords:** *Binning, duplicate detection, heterogeneous structure, network pruning, posterior probability, SVM, XML*

## **1. Introduction**

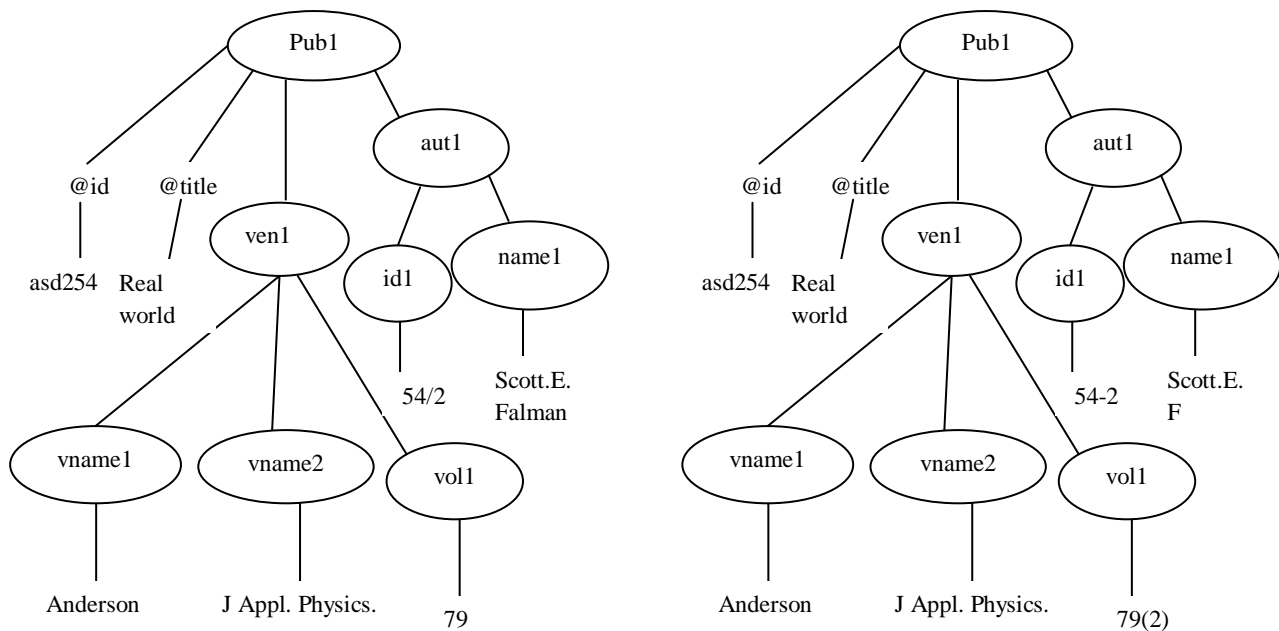
eXtensible Markup Language (XML) is widely used in most of the business applications to exchange data within the network and it is also used for publishing data on web. Most of the time XML data comes with errors and inconsistencies inherent in the real world data. It is necessary to ensure the quality of data published on web as it is created from distributed and heterogeneous data source. Data quality however can be compromised by different types of errors, which can have various origins [1].

Identifying and eliminating duplicates has become one of the challenging tasks as the data may not look exactly similar. Now a day's data's are given different representation as a result identifying different representation of the same entity turn out to be a problem in the field of duplicate detection. It is essential to use a correct matching strategy for identifying if they refer to the same real world entity or not.

Due to the extensive applications in various fields, duplicate detection has been studied profoundly for data stored in relational tables. While detecting duplicates in relational structure the tuples are compared and their similarity scores are computed based on their attribute values. In most cases, it omits some of the data's as foreign keys are used to connect tables. Many algorithms were developed which considered hierarchical and semi structured data [2] [3] [4] [5].

Both the figures represent the same publication details. The main difference between figures is the way in which they are represented. In both the figures nodes are labeled by their XML tag name. Leaf elements have text node which stores the actual data. Ven1 and aut1 are child nodes which in turn act as parent node for vname1, vname2, vol1 and name1, id respectively. The objective of duplicate detection is to efficiently prove that both the publication depicted in the figures in different format is identical despite of their structure.





**Figure1. XML element that represents the same publication detail in format.**

A hierarchical representation of XML data is shown in the Fig. 1. The elements are represented in different format and they differ from each other in the way they are structured.

**Contribution-**In this paper a supervised machine learning algorithm known as SVM is used for deriving conditional probabilities. A method known as binning is used to convert the output of SVM into an accurate posterior probability [6] [7]. To obtain a better result in terms of effectiveness a network pruning [8] is applied. Unlike other works the main aim is to improve the performance of detecting duplicates of differently structured hierarchical data. In this paper performance is evaluated by comparing the recall and precision values of XMLDup and Dogmatix with the proposed hybrid method using SVM.

**Layout -**This paper is organized as follows: Section 2 presents the related work. Section 3 summarizes the hybrid methodology. Section 4 describes the experimental setup and result analysis. Finally, section 5 concludes the work and presents a suggestion for future work.

## 2. Related Works

In this section various duplicate detection algorithms and techniques are explained.

Delphi [9] is used to identify duplicates in data warehouse which is hierarchically organized in a table. It doesn't compare all pairs of tuples in the hierarchy as it evaluates the outermost layer first and then proceeds to the innermost layer.

D. Milano et.al, [5] suggested a method for measuring the distance of each XML data with one another, known as structure aware XML distance. Using the edit distance measure, similarity measure can be evaluated. This method compares only a portion of XML data tree whose structure is similar nature.

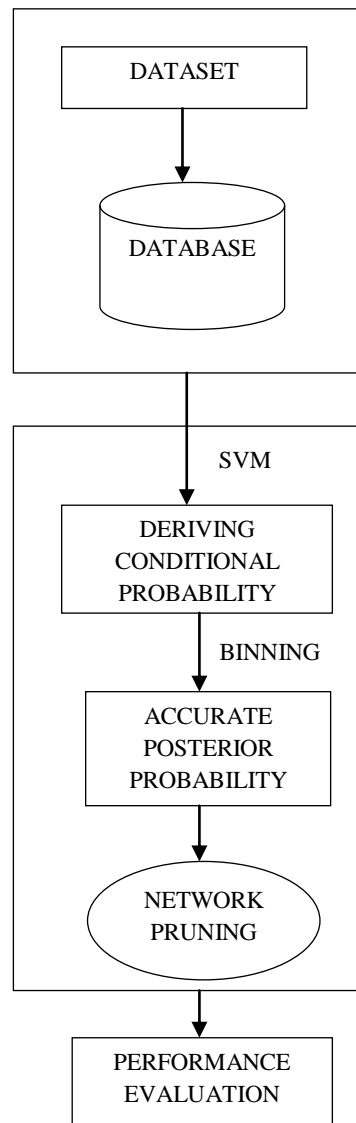
M. Weis et.al [2] proposed Dogmatix framework which comprises of three main steps: candidate definition, duplicate definition and duplicate detection. Dogmatix compares XML elements based on the similarity of their parents, children and structure. It also takes into account difference of the compared elements.

A novel method, XMLDup [3] is used for detecting fuzzy duplicates in hierarchical and semi structured XML data which considers the duplicate status of children and probability of the descendants being duplicates. Probabilities are derived efficiently using a Bayesian Network, which helps in determining the duplicates in XML data.

Network pruning [8] was the extension of XMLDup which was proposed to accelerate the Bayesian network evaluation time. In network pruning, a threshold is given and only those object pairs incapable of reaching the threshold was discarded. Proposed paper made a difference with this paper in the sense different structures are considered and conditional probability is estimated using SVM.

### 3. XML Duplicate Detection

The main goal of duplicate detection in XML data is to identify the XML element which represents the same real world entity. An XML document is considered as duplicates not only based on their structure but also on the way in which contents are represented. Each node is considered as duplicates based on their probability values. If estimated probability is above a given threshold value then it is considered as duplicate.



**Figure 2 Overview of duplicate detection**

Figure. 2 show the overview of detecting duplicates in XML data using the support vector machine classifier and network pruning.

#### 3.1 Deriving Conditional Probabilities Using SVM

A XML document is considered as duplicates based on the conditional probability. Using SVM probabilities of different structure can be calculated efficiently. While applying SVM, conditional probability is obtained as the output which is converted into an accurate posterior probability using binning [10], [11].

While using SVM, the dataset was divided into two: training and testing sets. SVMs learn a decision boundary between two classes by mapping the training examples onto a higher dimensional space and then determining the best separating hyper plane between those spaces [10]. Given a test example 'x', the SVM outputs a score that measures the distance of 'x' from the separating hyper plane. The sign of the score indicates

to which class 'j' example 'x' belongs, where  $j = \{1, 0\}$ . The obtained score has to be converted into an accurate posterior conditional probability. For the conversion purpose a histogram technique known as binning is used.

In the figures given above node pub1 is considered as duplicates only when its child node and value node are duplicates. A child node is considered as duplicates only when their value nodes are duplicates. In short a node is considered as duplicates only when all its child node and value nodes are duplicates.

Four conditional probabilities are stated below are from [8]:

Conditional Probability 1 (CP1): the probability of values of a node being duplicates depends on three factors: 1) all the individual value nodes are duplicates 2) all the individual value nodes are not duplicates and 3) some of the value nodes are duplicates based on the threshold value . From the examples if id, title, vname's, vol, name are duplicates then the pub1 nodes are duplicates otherwise if all the nodes are non-duplicates then pub1 is considered as non-duplicates. If only some nodes are duplicates, then duplicate probability is based on a given value 'a'.

Conditional Probability 2 (CP2): the probability of a children node being duplicates, given that each individual pair of children is duplicates. Aut1 and ven1 are duplicates only if their child nodes are duplicates.

Conditional Probability 3 (CP3): the probabilities of two nodes are duplicates given that their values and children are duplicates. Pub1 are duplicates only if both the children node ven1 and aut1 are duplicates and the value node id and title are duplicates.

Conditional Probability 4 (CP4): the probability of a set of nodes of the same type being duplicates given that each pair of individual nodes in the sets is duplicates.

Using SVM probability of a node being duplicate or non – duplicate can be determined easily. SVM first evaluate the probability of the parent node being duplicates. If there is a probability for parent node to be duplicates then the corresponding probability of the child node and value node being duplicates are evaluated. If there is probability of the child and value node to be duplicates then then the parent node is considered as duplicates.

After all the conditional probabilities have been derived binning is used to convert the conditional probability to accurate posterior conditional probability. In binning method the training instances are first ranked according to their scores. Then the process continues by dividing the instance into 'n' subsets of equal size. The corresponding estimated probability  $P(j|x)$  is the portion of training instances that actually fit to the class that has been predicted for the test example [6].

### 3.2 Network Pruning

An algorithm proposed in [8] is used to prune the non-duplicate node that doesn't cross the threshold value. The algorithm takes a node N and if the node probability score falls above a given threshold value then it is considered as duplicates otherwise that node is discarded. Higher the similarity score there are more chances of missing out the duplicate pairs. By lowering the similarity score network can be evaluated faster as there are more chances of crossing the threshold value easily. Even though network can be evaluated faster there are chances of missing out duplicates which can be considered as a disadvantage.

## 4. Experimental Setup and Result Discussion

This section gives a brief description about the experiments performed using the dataset. The same dataset which was used [8] are taken for the process of duplicate detection. The experimental evaluation was performed on Cora dataset.

Experiments were performed to compare the effectiveness of the hybrid method using SVM with XMLDup and Dogmatix. To measure the effectiveness of the proposed method two parameters are used: recall and precision.

$$\text{Precision} = \frac{tp}{(tp+fp)} \quad (4.1)$$

$$\text{Recall} = \frac{tp}{(tp+fn)} \quad (4.2)$$

where, tp(true positive ) is correctly identified duplicates, fp(false positive) is number of non-duplicate nodes which are identified as duplicates and fn(false negative) are the number of duplicates nodes identified as non-duplicates.

The experimental assessment was performed on an intel core i5 CPU at 2.67GHz with a 4Gb RAM. It is fully implemented in Microsoft Visual Studio.

To measure the effectiveness of the proposed method it is compared with XMLDup and Dogmatix.

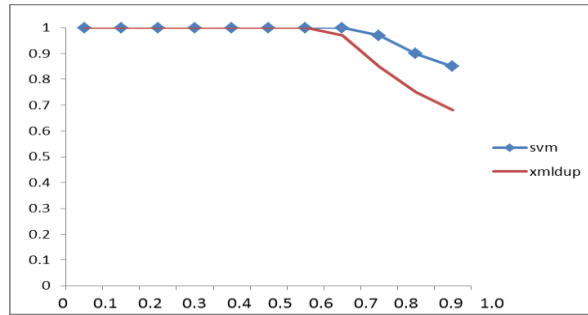


Figure 3 Comparison of results of XMLDup and hybrid method using SVM using Cora dataset.

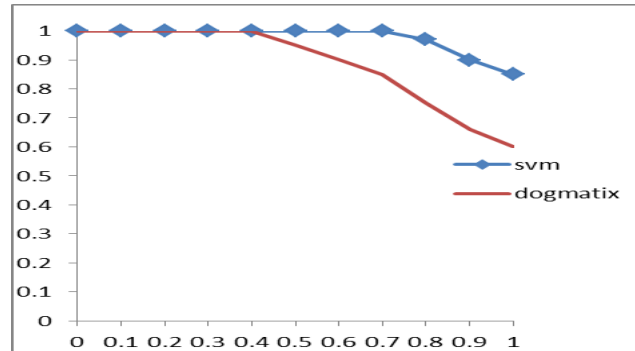


Figure 4 Comparison of results of Dogmatix and hybrid method using SVM using Cora dataset.

Figure 3 and Figure 4 represents the comparison result for XMLDup, Dogmatix and proposed method using SVM representing the recall and precision values for the dataset Cora. The proposed approach shows a better result compared to other algorithms.

Table 1 Recall and precision values for various pruning factor

Pruning factor	SVM		XMLDUP		DOGMATIX	
	r	p	r	p	r	p
0.4	75	99	98	83	92	73
0.5	75	99	98	83	92	73
0.6	75	99	99	87	93	81
0.7	82	100	99	87	93	81
0.8	82	100	99	95	95	83
0.9	82	100	99	95	95	83
1.0	82	100	99	95	95	83

The table which is shown above gives a detailed view about the performance of various algorithms based on different pruning factor. While using SVM it shows a high value of precision compared to other algorithms even though the recall value is low.

### 5. Conclusion

In this paper, a machine learning algorithm known as SVM is proposed for deriving conditional probabilities for the detection of duplicates and a technique known as binning is used to convert the output of SVM to an accurate posterior probability. Estimating the probability using SVM increases the rate of duplicate detection. SVM not only consider contents but it also takes into account XML objects with different structures. The proposed method achieves an improvement in the value of precision on different structured data.

### References

- [1] E. Rahm and H.H. Do, "Data Cleaning: Problems and Current Approaches," IEEE Data Eng. Bull., vol. 23, no. 4, pp. 3-13, Dec.2000.
- [2] M. Weis and F. Naumann, "Dogmatix Tracks Down Duplicates in XML," Proc. ACM SIGMOD Conf. Management of Data, pp. 431-442, 2005.
- [3] L. Leitao, P. Calado, and M. Weis, "Structure-Based Inference of XML Similarity for Fuzzy Duplicate Detection," Proc. 16th ACM International Conf. Information and Knowledge Management, pp. 293-302,2007.
- [4] A.M. Kade and C.A. Heuser, "Matching XML Documents in Highly Dynamic Applications," Proc. ACM Symp. Document Eng. (DocEng), pp. 191-198, 2008.
- [5] D. Milano, M. Scannapieco, and T. Catarci, "Structure Aware XML Object Identification," Proc. VLDB Workshop Clean Databases (CleanDB), 2006.
- [6] J. Drish "Obtaining calibrated probability estimates from support vector classifiers: project proposal"
- [7] B. Zadrozny and C. Elkan. "Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers". Proceedings of the Eighteenth International Conference on Machine Learning, 2001.
- [8] M. Weis, L. Leitao and P. Calado "Efficient and Effective Duplicate Detection in Hierarchical Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 25, May 2013.
- [9] R. Ananthakrishna, S. Chaudhuri, and V. Ganti. "Eliminating fuzzy duplicates in data warehouses," In International Conference on Very Large Databases, Hong Kong, China, 2002.
- [10] J. Drish "Obtaining calibrated probability estimates from support vector classifiers"
- [11] J.T. Kwok "Moderating the Outputs of Support Vector Machine Classifiers". In IEEE -NN,1995

# Survey Paper on Integrity Auditing of Storage

Ugale Santosh A<sup>1</sup>

1M.E. Computer AVCOE, Sangmner, India

## ABSTRACT:

Cloud servers is a model for enabling convenient, on-demand network access to a shared pool of configurable server resources (networks, memory, storage, cpu, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud service provider interaction. The Cloud servers models offers the promise of massive cost savings combined with increased IT agility due to pay per consume. However, this technology challenges many traditional approaches to hosting provider and enterprise application design and management. Cloud servers are currently being used; however, data security cited as major barriers to adoption in cloud storage. Users can store data and used on demand or for the applications without keeping any local copy of the data. Users can able to upload data on cloud storage without worrying about to check or verify integrity. Hence integrity auditing for cloud data is more important task to ensure users data integrity. To do this user can resort the TPA (Third Party Auditor) to check the data on the cloud storage. TPA is the expertise and having knowledge and capabilities which users can unable to check. TPA audit the integrity of all files stored on the cloud storage on behalf of users and inform the results. Users should consider auditing process will not cause new vulnerability against the users data also ensures integrity auditing will not cause any resources problem.

**Keywords:** Auditing, Cloud, Cloud servers, Data integrity, Data privacy, Security, Storage

## I. INTRODUCTION

Integrity auditing is something you need to have on cloud storage. Different threats imagine a hacker placing a backdoor on storage using applications; modify files, change permissions, or changing your order form to email him a copy of everyone's credit card and other information while leaving it appear to be functionally normally without any problem. By auditing process and setting up convenient period scan reporting, this notifies user within hours of when any file was changed, modified, added or removed. It also helps establish an audit trail in the event cloud storage is compromised. Cloud servers has been envisioned as the next-generation information technology (IT) architecture for government, research, and industry, due to configurable server resources and long list of advantages: on-demand self-service, dynamic resources allocation, Auto-Scaling technology, fast, secure, ubiquitous network access, location independent, resource elasticity, pay per consume, higher uptime and transference of risk [14].

Cloud Computing is remodeling the very nature of how businesses use information technology. One elementary side of this paradigm shifting is that data is being centralized or outsourced to the Cloud server. From users' perspective, including both user and enterprises, uploading data to the cloud server in a flexible on-demand manner brings appealing benefits: relief of the burden for storage and security management, global data access with independent geographical locations, and saving of capital expenditure on security [13], hardware resources and maintenance, etc. whereas Cloud storage makes these features more appealing than ever, it also brings new security vulnerability towards users' data. As a result, the integrity of the data in the cloud is being put at risk due to the different reasons. Although the infrastructures under the cloud provider are much more powerful and secure than local computing devices, they are still facing the different internal and external threats for data integrity. Secondly, there do exist various motivations for hosting provider to behave unfaithfully towards the cloud users regarding the status of their remote data. In short, although outsourcing data to the cloud servers is economically attractive for long-term huge data storage, cloud service provider does not provide any guarantee on data integrity and security. This drawback, if not properly addressed, could impede the successful deployment of the cloud server's design. As users data on remote storage, traditional cryptographic primitives for the purpose of data security protection cannot be adopted [10] directly specifically, downloading data on native system for its integrity verification is not a practical solution due to the transmission cost across the network and security reasons. Considering the large size of the outsourced data store and the user's constrained resources capability, the work of auditing the data correctness in a cloud server environment can be expensive for the cloud server users [7], [9]. Moreover, the overhead of using cloud server storage should be minimized as much as possible, such that cloud user does not need to perform huge operations to use the cloud server data. For example, it is desirable that cloud users do not need to worry about the need to verify the integrity of the data before or after the data retrieval. Besides, there may be multiple user's accesses the same cloud storage for different purpose and applications, say in an enterprise setting.

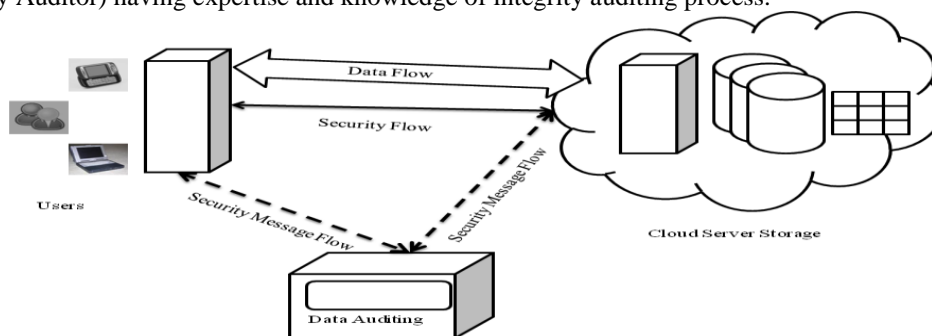
To make it ensure the data integrity and minimize the cloud server computation resources as well as online burden on cloud users', it is of critical importance to enable public auditing process for cloud data storage, so that cloud users may resort to an independent third party auditor (TPA) to audit the data stored on the cloud storage whenever necessary. The TPA, who has knowledge and capabilities that users do not, can check the data integrity of all the data stored in the cloud periodically on behalf of the cloud users, which provides a much more easier and affordable way for the users to ensure their cloud data storage integrity. Moreover, in addition to help users to evaluate the risk of their subscribed cloud data services, the audit result obtained from TPA would also be beneficial for the CSP or hosting provider to improve their security related to storage platform. In a word, auditing services will play an important role for this cloud economy to become fully established; where users will need ways to assess risk and gain trust in the cloud service provider or cloud storage. Currently, the notion of public auditability has been proposed in the context of ensuring remotely stored data integrity under different system and security models [8], [9], [10], [11], [12].

Auditability process allows a third party, in addition to the user himself, to verify the integrity of remotely stored cloud data. However, most of these schemes [8], [9], [11] do not consider the privacy protection of users' data against external auditors. Indeed, TPA may potentially reveal user data information to the auditors. This severe drawback greatly affects the security of these protocols in Cloud storage. From the perspective of protecting data privacy and integrity, the users, who own the data on cloud server and rely on TPA auditing process just for the storage security and integrity of their data, do not want TPA auditing process introducing new vulnerabilities of unauthorized data leakage towards their data security [12].

Also there are some legal regulations on outsourced data that is, data not to be leaked to external parties. Without properly designed auditing protocols, encryption itself cannot prevent data from "flowing away" towards TPA during the public auditing process. The reason, it does not completely solve the problem of protecting data privacy from external parties but just reduces it to the key management. Vulnerability of unauthorized data leakage still remains a problem due to the potential exposure of decryption keys. Therefore, how to enable an auditing protocol keeping data private, independent to data encryption is the problem I am going to tackle in this paper.

## II. PROBLEM STATEMENT

The system model I have considered cloud data storage or files storage involving three different entities. As illustrated in figure 1 [1], the cloud users who store the huge amounts of data in the form of files on the cloud storage. Files may be in different types such as binary files, data files, logs files, hidden files. The cloud servers, which fully managed by the hosting or cloud service provider for the data storage space and different resources like network connection, backup facilities and different level security. Third entity is TPA (Third Party Auditor) having expertise and knowledge of integrity auditing process.



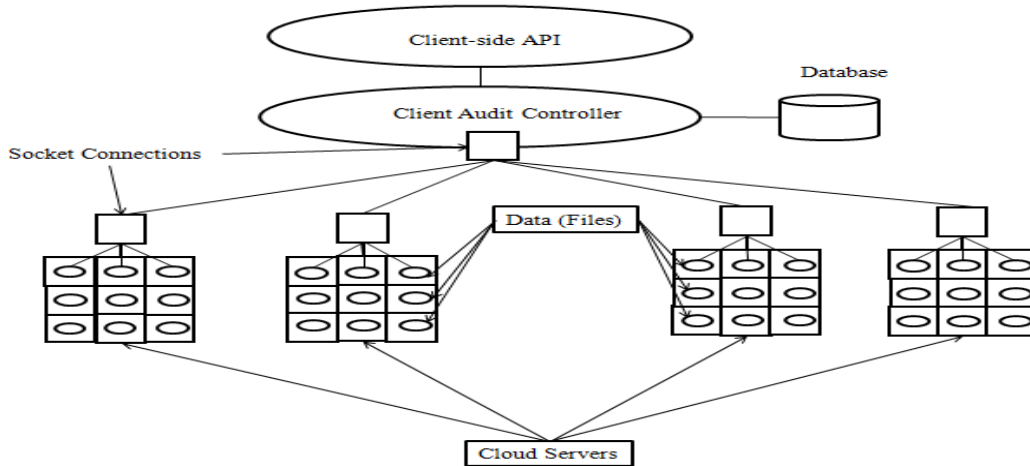
"Figure 1. Cloud architecture"

Cloud service provider is responsible for storage management, maintenance, scalable, pay per consume, location independent, higher availability and low cost data storage. Users upload and download data dynamically from storage space on the cloud server for its own application purpose. Users always need to ensure, data stored on the server is correct and maintained properly. To avoid computational resources and ensure data integrity and security users resort to TPA to audit the data on behalf of user on cloud server.

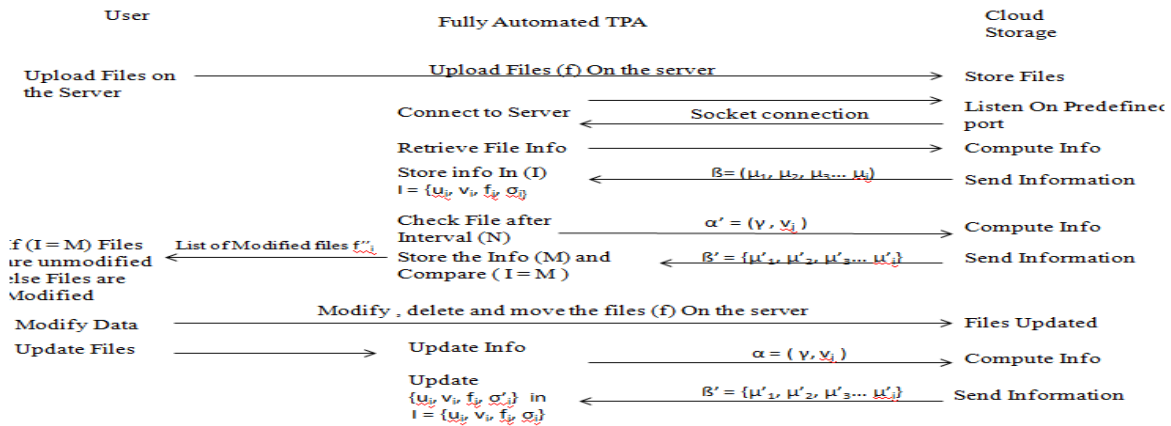
User's data could be hack, modified or changed by internal or external entities. It may includes software bugs, backdoors in different applications, outdated applications versions, plug-in, themes, templates, bugs in system or economically motivated hackers, malicious code and different upload forms. Cloud servers always provide better security but due to different integrity threats towards data like vulnerable functions used in application, outdated applications versions, plugins, themes, templates, bugs in system backdoors in application, applications from the un trusted sources which come with preloaded outdoors, hardware failure, network issue there is changes of data loss. Cloud service provider always try to hide these details from users to their own benefits as well as maintain industry reputation the reason cloud users cannot completely trust on the cloud service provider. With the help of auditing procedure user can gain trust as well as audit his data more efficiently.

### III. PROPOSED WORK

This section presents integrity auditing scheme which provides a complete outsourcing solution of data. After introducing notations considered and brief preliminaries, I have started from an overview of proposed Integrity auditing scheme. Then, I am presenting main scheme and show how to extent proposed scheme to support integrity auditing for the TPA upon delegations from multiple users. Finally, I have proposed how to generalize integrity auditing keeping data privacy scheme and its support of dynamic data. Figure 2 illustrate the overview of integrity auditing structure.



“Figure 2. Integrity auditing block diagram”



“Figure 3. Auditing protocol”



#### IV. IMPLEMENTATION DETAILS

##### 4.1 Mathematical Model

$S = \{x, e, i, o, f, DD, NDD, success, failure\}$

Let  $S$  be the solution perspective of the class

$x =$  Initial state of the class Initialize ()

$x = \{Initialize ()\}$  sets the default values for all variables.

Input  $i = (I1, I2)$

$I1 = \{\{U\}\{V\}\{F\}\{\sigma\}\}$

DD=deterministic data it helps identifying the load store functions or assignment functions.

NDD=Non deterministic data of the system  $S$  to be solved.

Success-desired outcome generated.

Failure-Desired outcome not generated or forced exit due to system error.

Set of 'k' cloud Users  $U = \{u_1, u_2, u_3, \dots, u_k\}$

Set of 'm' cloud servers  $V = \{v_1, v_2, v_3, \dots, v_m\}$

Set of files on cloud storage  $F = \{f_1, f_2, f_3, \dots, f_n\}$

Set of file tags  $\sigma_i = \{f+p+n+u+g+s+acl+b+selinux+md5+sha256\}$ ,  $i \in (1, n)$

$p =$  File permissions,  $t =$  File type,  $i =$  File Inode number

$u =$  File User ID,  $g =$  File Group ID,  $s =$  File Size

$b =$  File Block count,  $m =$  File Modified time

$a =$  File Access Time (when the file was last read,  $c =$  is the inode change time,  $n =$  Number of links For file

$S =$  Check for growing size,  $md5 =$  md5 hash,  $sha1 =$  sha1 hash,  $f =$  File name,  $I =$  Initial Values in

Database,  $N =$  Interval of auditing process,  $M =$  New Value database,  $LI =$  List Of files,  $ST =$  Detail info of

modified files, Set of file tags  $\sigma$  calculated based on the file types,  $\gamma =$  directory path,  $\alpha =$  query  $v =$  cloud IP

address,  $\beta =$  set of results  $\mu =$  consist of file stats.

[ Data DATA =  $f+p+n+u+g+s+acl+b+selinux+md5+sha256$ ]

[ Growing files GROW =  $p+u+g+i+n+S+acl+selinux$  ] [ Password and shadow files IMP =  $A+sha256$  ]

[ Binary and Configuration files. FIXF =  $A+sha256$  ] [ Hidden file PERM =  $p+u+g+i+acl+selinux$  ]

[ Directories DIR =  $p+n+i+u+g+acl+selinux$  ]

Where  $A = p + n + i + u + g + b + s + m + c + acl + selinux + md5$   $H = sha1 + sha256 + sha512$

##### 4.1.1 Initialize ()

TPA send Query initialize()

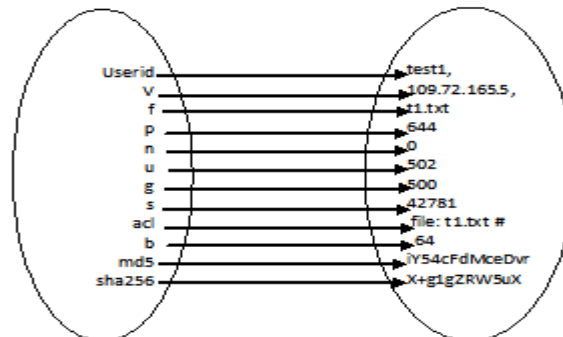
$\alpha = (\gamma, v_j)$  where,  $\gamma \in n$  and  $v_j$  is a  $j^{th}$  cloud server.

( $\gamma$  is set or path of (n) files and  $v_j$  is cloud IP address)

$v_j$  cloud server produces  $\beta = (\mu_1, \mu_2, \mu_3 \dots \mu_i)$

Where,  $\mu_i$  comes from  $(f_1, f_2, f_3 \dots f_n)$  consists of pair  $(f_i, \sigma_i)$ . TPA store the received values in database (I)

Figure 4 show sets of variables and values.  $I = \{u_i, v_j, f_i, \sigma_i\}$  Where,  $u_i$  is user,  $v_j$  cloud server and  $\sigma_i$  consist of signature tag of file  $f_i$



“Figure 4. Sets of variables”

$f = \{update (), Check integrity ()\}$

### 4.1.2 Update()

A step after user uploads/modified the files on cloud server.

TPA send Query Update  $\alpha = (\gamma, v_j)$  where  $\gamma \in n'$  and  $v_j$  is a  $j^{\text{th}}$  cloud server.  $n'$  updated files.

Set of tags  $\sigma'_i = \{f+p+n+u+g+s+acl+b+selinux+md5+sha256\}$ ,

$i \in (1, n')$  where  $\sigma'$  updated files tags Number of files  $F = \{f_1, f_2, f_3, f_4, \dots, f_n\}$

Cloud server produces  $\beta' = \{\mu'_1, \mu'_2, \mu'_3, \dots, \mu'_i\}$  Where  $\mu_i$  comes from  $(f_1, f_2, f_3, \dots, f_n)$  consists of pair  $(f_i, \sigma'_i)$

TPA add/replace the  $\beta'$  values  $\{u_i, v_i, f_i, \sigma'_i\}$  in  $I = \{u_i, v_i, f_i, \sigma_i\}$

$I = \{u_i, v_i, f_i, \sigma_i\}$  where  $u_i$  is user,  $v_i$  cloud server and  $\sigma_i$  consist of signature tag of file  $f_i$

### 4.1.3 Check integrity()

Initial values  $I = \{u_i, v_i, f_i, \sigma_i\}$  where,  $u_i$  user,  $v_i$  cloud sever IP,  $\mu_i = (f_i, \sigma_i)$  file name with file stats.

Interval to check integrity (N)

Set of tags  $\sigma'_i = \{f+p+n+u+g+s+acl+b+selinux+md5+sha256\}$ ,  $i \in (1, n')$  where  $\sigma'$  updated files tags

Number of files  $F = \{f_1, f_2, f_3, f_4, \dots, f_n\}$

TPA to cloud server Query Check  $\alpha' = (\gamma, v_j)$

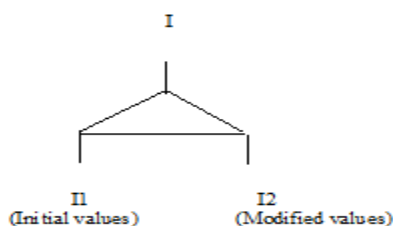
Produces  $\beta' = \{\mu'_1, \mu'_2, \mu'_3, \dots, \mu'_i\}$  where  $\mu_i$  comes from  $(f'_1, f'_2, f'_3, \dots, f'_n)$

TPA store the received  $\beta'$  values  $\{f'_i, \sigma'_i\}$  in database (M) along with user and server details.

$M = \{u_i, v_i, f'_i, \sigma'_i\}$

TPA Search M  $\{u_i, v_i, f'_i, \sigma'_i\}$  in to the database I  $\{u_i, v_i, f_i, \sigma_i\}$

If  $M \{u_i, v_i, f'_i, \sigma'_i\} \in I \{u_i, v_i, f_i, \sigma_i\}$



“Figure 5. Results comparison”

As per the Figure 5 TPA system compares the values

Success- If  $M \{u_i, v_i, f'_i, \sigma'_i\} \neq$  Search result (I)  $\{u_i, v_i, f_i, \sigma_i\}$

**Results::** Files modified lists  $(f'_i)$  Else  $M \{u_i, v_i, f'_i, \sigma'_i\} =$  Search result (I)  $\{u_i, v_i, f_i, \sigma_i\}$

**Results::** Files not modified

Failure-Desired results are not generated.

In this scheme, work based on the six phases includes Install client, connect, upload, initialize, check/compare and update

## REFERENCES

- [1] Cong Wang, Sherman S.M Chow, Qian Wang, Kui Ren and Wening Lou, “Privacy-Preserving Public Auditing for Secure cloud storage” in IEEE transaction on computers vol 62 No 2 February 2013.
- [2] Cong Wang, Qian Wang, Kui Ren, Ning Cao, and Wenjing Lou “Toward Secure and Dependable Storage Services in Cloud Computing” IEEE Transaction on Services Computing vol 5 No 2 April-June 2012.
- [3] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou And Jin Li “Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing” IEEE transaction Paper on Parallel and Distributed Systems vol 22 No 5, pp. 847-859, May 2011.
- [4] Kan Yang and Xiaohua Jia “An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing” IEEE transaction on parallel distributed system, Vol 24 No 9 September 2013.
- [5] Yan Zhu, Hongxin Hu, Gail-Joon Ahn and mengyang Yu “Cooperative Provable Data possession for Integrity Verification in Multicloud Storage.” IEEE Transactions on parallel and distributed system, Vol 23, No. 12, pp. 2231-2244, December 2012.
- [6] Shucheng Yu, C. Wang, K. Ren, and Wenjing Lou, “Achieving secure, scalable, and fine-grained access control in cloud computing,” in *Proc. of IEEE NFOCOM'10*, San Diego, CA, USA, March 2010.
- [7] Cloud Security Alliance, “Security guidance for critical areas of focus in cloud computing,” 2009, <http://www.cloudsecurityalliance.org>.
- [8] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable data possession at untrusted Stores,” in *Proc. of CCS'07*, Alexandria, VA, October 2007, pp. 598–609.
- [9] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, “Enabling public verifiability and data dynamics for storage security in cloud computing,” in *Proc. of ESORICS'09, volume 5789 of LNCS*. Springer-Verlag, Sep. 2009, pp. 355–370.
- [10] A. Juels and J. Burton S. Kaliski, “Pors: Proofs of retrievability or large files,” in *Proc. of CCS'07*, Alexandria, VA, October 2007, pp. 584–597.
- [11] H. Shacham and B. Waters, “Compact proofs of retrievability,” in *Proc. of Asiacrypt 2008*, vol. 5350, Dec 2008, pp. 90–107.
- [12] M. A. Shah, M. Baker, J. C. Mogul, and R. Swaminathan, “Auditing to keep online storage services honest,” in *Proc. Of HotOS'07*. Berkeley, CA, USA: USENIX Association, 2007, pp.1–6.
- [13] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the clouds: A berkeley view of cloud computing,” University of California, Berkeley, Tech. Rep.
- [14] P. Mell and T. Grance, “Draft NIST working definition of cloud computing,” Referenced on June. 3rd, 2009 Online at <http://csrc.nist.gov/groups/SNS/cloud-computing/index.html>, 2009.

# Lookup table embedded in (FPGA) for network security

ZuhirNemerAlaaraj<sup>1</sup>, Abdelrasoul Jabar Alzubaidi<sup>2</sup>

1 Sudan Academy of Sciences (SAS); Council of Engineering Researches & Industrial Technologies

2 Electronic Dept. - Engineering College –Sudan University for science and Technology

## ABSTRACT

*This work proposes a solution to improve the security of data through flexible bitstream encryption, by using the lookup table (LUT) that is embedded in the FPGA.*

*This technique concentrates on building high data security to make it difficult for an adversary to capture the real data.*

*The syntheses proposed can be implemented in two steps:*

*First step: programming the FPGA to create a LUT by VHDL language the LUT has a predefined length and contents.*

*Second step: applying security strategy.*

*A high security and more reliability can be achieved by using this mechanism when applied on data communication and information transformed between networks.*

**KEY WORDS:** LUT, FPGA.VHDL, embedded system, security.

## I. INTRODUCTION:

The FPGA architecture consists of three types of configurable elements - a perimeter of input/output blocks (IOBs), a core array of configurable logic blocks (CLBs), and resources for interconnection. The IOBs provide a programmable interface between the internal arrays of logic blocks (CLBs) and the device's external package pins, see figure (1). CLBs perform user-specified logic functions, and the interconnect resources carry signals among the blocks [1].

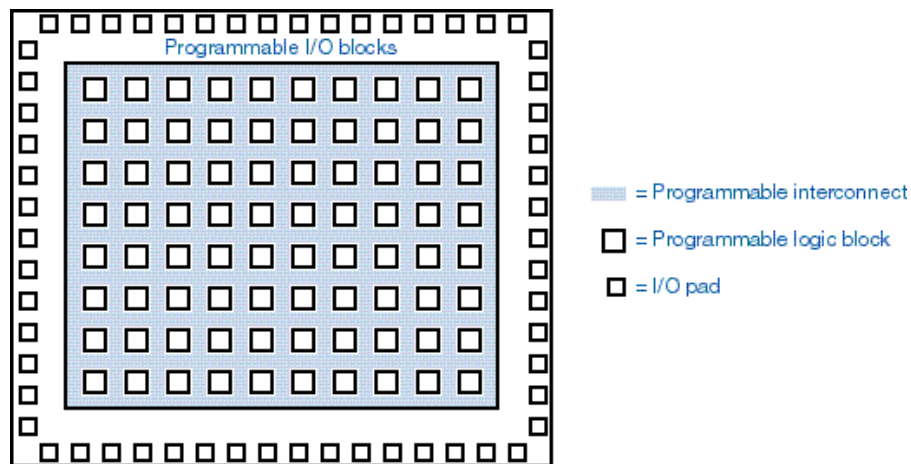


Figure 1 Structure of FPGA

Those FPGAs are attractive for executing the actual cryptographic algorithms and are, thus, of particular importance from a security point of view.

FPGA produced by XILINX will be used.

Today FPGAs represent an efficient design solution for numerous systems. They become necessary to improve data security.

## II. SYNTHESIS:

All randomly stored values in FPGA are indexed with the addresses at the lookup table. These random values are stored by the user (program). The obtained structures from this mechanism are blind, and the adversary who doesn't know the algorithms cannot reverse it to the original feature values. These types of techniques were developed to prevent an attacker from knowing the real data. The logic design of a module can be done with a standard hardware description language, such as Verilog or VHDL.

## III. DESIGN THEORY:

The procedure includes embedding a LUT to deploy security. In the beginning we start by defining two variables which we apply on the lookup table:

1. The length of lookup table which is the number of data bytes we would like to embed in it.
2. The content of data inserted by the programmer in the LUT.

Here, when we deploy security strategy every byte from the input data stream is altered by XORing it with the written value in the LUT. The process continues till the end of the lookup table content is finished. The operation is repeated if the data stream is present. Hence, we obtain a cipher text which is ciphered by the lookup table content XORed with data, which is known only to the programmer. The attacker cannot retrieve the original data because he can't know two very important factors in ciphering:

1. The lookup table length and so the number of repeats.
2. The lookup table content which was written by the programmer.

In this methodology the programmer can change the security strategy at his well.

## IV. METHODOLOGY

### Step1:

#### *Programming of the FPGA:*

The process of implementing a design on an FPGA can be broken down into. The VHDL code is converted into device netlist format. Then the resulting file is converted into a hexadecimal bit-stream file, or bit file. This step is necessary to change the list of required devices and interconnects into hexadecimal bits to download to the FPGA. The bit file is downloaded to the physical FPGA. This final step completes the FPGA synthesis procedure by downloading the design onto the physical FPGA see figure (2).

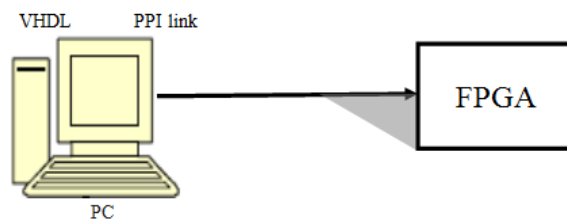


Figure 2 programming the FPGA

### Step2:

#### *Application for security on the network*

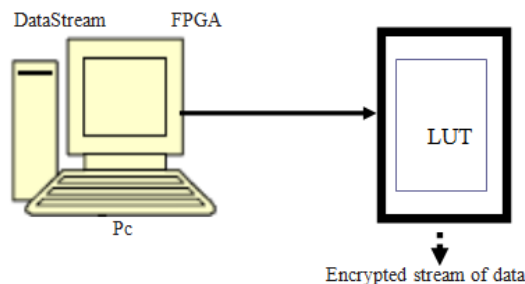


Figure 3 security implementation

**V. THE ALGORITHM:**

The design proposes a creation of ten bytes lookup table.  
 The contents of the ten bytes can have any value.  
 The flow chart for the design is shown in figure (4).

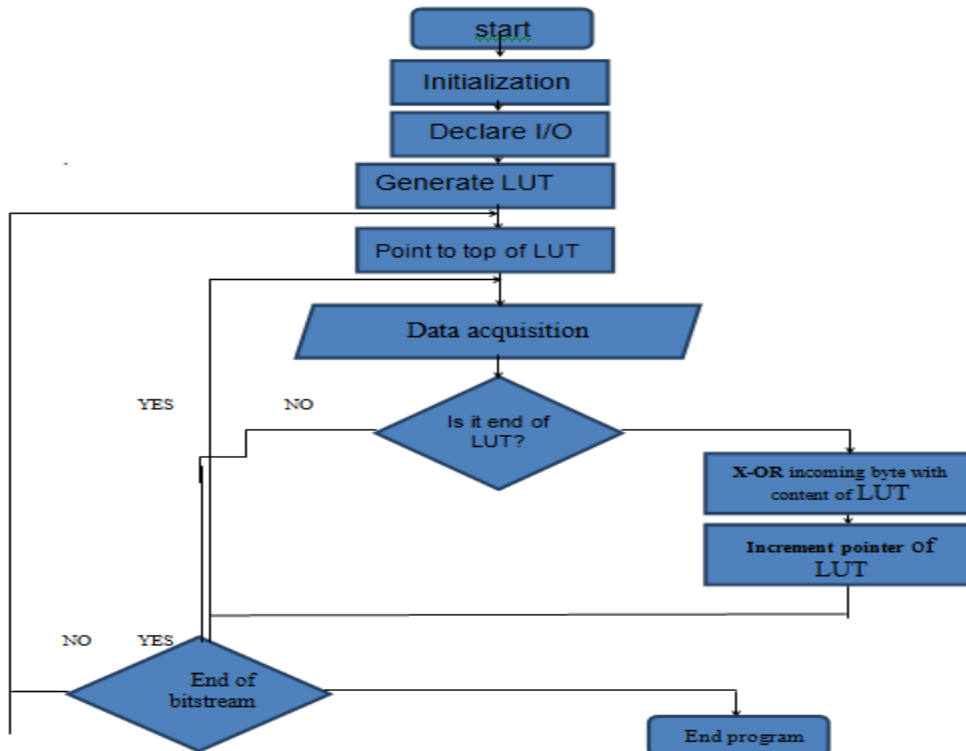


Figure 4 the flow chart

**VI. RESULTS:**

Assume a data stream being input to the FPGA (column 1 in figure 5), and the ten characters of the lookup table (column 2 in figure 5).

Xoring column 1 with column 2 gives us encrypted data output shown in column 3 of figures (5).

Figure (5) below shows the complete picture of the assumption.

The data stream and the contents of the (LUT) are given in (ASCII) code.

Column 1	Column 2	Column 3
<b>Data stream (ST)</b>	<b>LUT</b>	<b>ST(XOR)LUT</b>
(41)H	(45)H	(04)H
(43)H	(4C)H	(0F)H
(45)H	(45)H	(00)H
(47)H	(43)H	(04)H
(49)H	(54)H	(1D)H
(4B)H	(52)H	(19)H
(4D)H	(4F)H	(02)H
(4F)H	(4E)H	(01)H
(51)H	(49)H	(18)H
(53)H	(43)H	(10)H

Figure 5 example based on LUT technique

## **VII. SUMMARY AND CONCLUSION:**

A high security and more reliability and high performance can be achieved by using this mechanism, when applied on data communication and information transferred between networks.

This paper is demonstrating new security structures concepts embedded in self reconfigurable VLSI technology environment. The resulting secret ciphers exhibit new security application horizons due to the particular possibility of constructing autonomous practical secret unknown functions. Keeping functions secret was assumed as a non-realistic assumption in cryptographic systems [2].

## **REFERENCES**

- [1.] A. S. Daniel Ziener and T. Jürgen. Identifying FPGA IP-Cores based on lookup table content analysis. In Field Programmable Logic and Applications, August 2006. <http://www12.informatik.uni-erlangen.de/publications/pub2006/zienerfp106.pdf>
- [2.] D. Agrawal, B. Archambeault, J. R. Rao, and P. Rohatgi. The EM side-channel(s). In Cryptographic Hardware and Embedded Systems Workshop, volume 2523 of LNCS, August 2002. <http://www.springerlink.com/content/mvtxbq9qa287g7c6/>

# Survey of Steganalysis Technique for Detection of Hidden Messages

Vanita J. Dighe,

AVCOE, Sangamner, A'nagar, MS, India.  
Email:vanitadighe1@gmail.com.

Prof. Baisa L. Gunjal,

Assistant Professor, AVCOE, Sangamner, A'nagar, MS, India.  
Email:hello\_baisa@yahoo.com.

## ABSTRACT

*Still and multi-media images are subject to transformations for image compression and steganographic hiding and digital watermarking. Here new measures and techniques for detection and analysis of steganographic embedded content. We show that both statistical and pattern classification techniques using our measures provide reasonable discrimination schemes for detecting embeddings of different levels. These measures are based on a few statistical properties of bit strings and wavelet coefficients of image pixels. There are Techniques for information hiding known as steganography are becoming increasingly more popular and spread over a large area.*

*The purpose of steganography is to send secret messages after embedding them into public digital multimedia. It is preferred to hide as many messages as possible per change of the cover-object. In general, for given messages and covers, the steganography that introduces fewer embedding changes will be less detectable, i.e., more secure. Two fields of study have been projected to develop the communication security: cryptography and information hiding. Although they are both applied to the protection of secret message, the major difference is the appearance of the transmitted data.*

**INDEX TERMS:** *Steganography, Least significant bit (LSB), Exploiting Modification Direction (EMD), Diamond Encoding (DE), Optimal Pixel Adjustment Process (OPAP), Pixel Pair Matching (PPM), Adaptive Pixel Pair Matching (APPM).*

## I. INTRODUCTION

All natural images have a lot of correlation among neighboring pixels. Image pixel data has statistical functionalities. All these are disturbed by the process of embedding. These are exploited in steganalysis of images. Data concealment may be a technique that conceals information into a carrier for transference secret messages confidentially [2], [3]. Digital pictures are wide transmitted over the Internet; thus, they usually functions a carrier for covert conversation. Pictures used for carrying information are termed as cover pictures and pictures with information embedded are termed as stego pictures. Once embedding, pixels of cover pictures are going to be changed and distortion occurs. The distortion caused by information embedding is named the embedding distortion [4]. A decent data-hiding methodology must be capable of evading visual and applied detection [5] whereas providing an adjustable payload [6]. The least vital bit substitution methodology, spoken as LSB during this study, may be a well-known data-hiding methodology. This method is easy to implement and is one of the favored hiding techniques. However, in LSB hiding, the pixels with even values are going to be accumulated by one or unchanged. The pixels with odd values are going to be decreased by one or unchanged. Therefore, the unbalanced embedding distortion emerges and is at risk of steganalysis [7], [8].

This technique embeds fixed-length secret bits into the least significant bits of pixels by directly replacing the LSBs of the cover image with the secret message bits. This technique is simple which typically effects noticeable distortion once the quantity of hidden bits for each pixel exceeds three. Many methods have been proposed to reduce the distortion induced by LSBs substitution. In 2004, Chan et al [9], Proposed a simple

and efficient optimal pixel adjustment process (OPAP) method to reduce the distortion caused by LSB replacement. The LSB and OPAP methods employ one pixel as an embedding unit, and hide data into the rightmost  $r$  LSBs. Other group of data-hiding methods employs two pixels as an embedding unit to hides a message digit notational system and termed as data-hiding methods known as pixel pair matching (PPM). Here we study a new data hiding method to reduce the embedding impact by providing a simple extraction function and a more compact neighborhood set. The given technique embeds more messages and thus increases the hiding efficiency. The image quality obtained by the given technique not only performs better than those obtained by (OPAP) and Diamond Encoding (DE), but also brings higher payload with less detect ability.

## II. BACKGROUND

The approach of embedding information into a picture, image steganography techniques may be divided into the subsequent groups: Spatial Domain or Image Domain and Transform Domain or Frequency Domain.

### 2.1 Spatial domain techniques:-

Operate directly on pixels. Spatial domain reversible information activity is performed supported the ways Difference Expansion (DE) and Histogram modification. The previous methodology provides higher capability whereas the later provides higher quality image. In Difference Expansion (DE), the embedded bit stream includes a pair of elements. The primary half is that the payload that conveys the secret message and therefore the second half is that the auxiliary data that contains embedding data. The dimensions of the second half must be kept little to increase embedding capability.

LSB embedding is un-compressed pictures is of most interest. Hence, we have a tendency to concentrate fully on LSB information hiding and detection for uncompressed pictures. There are measure of few schemes that is slight variants of LSB replacement techniques. There are few techniques which are slight variants of LSB replacement techniques. Instead of replacing Least Significant Bit's of pixel value, the pixel value is incremented or decremented depending upon bit of the data and value of the pixel. Embedding information into least significant bit won't be seen by the human eye. Therefore the stego image appears like original or cover image.

The drawback of LSB is though it is simplest and easiest way for embedding information into pictures, once more data is hidden, the image loses its quality. Statistical analysis of the stego image ends up in the suspicion of hidden data.

### 2.2 Transform domain techniques:-

Images are first transformed and then the message is hidden into it. These are robust methods for data embedding. It is more complex method to hide secret message into an image. It performs data hiding by manipulating mathematical functions and transformations of image. Transformation of cover image is performed by tweaking the coefficients and inverts the transformation. Popular transformations include the two-dimensional discrete cosine transformation (DCT) [15]) discrete Fourier transformation (DFT) and discrete wavelet transformation (DWT) [14]) that are commonly used in image steganalysis.

It is accepted from Fourier theory that a signal will be expressed because the total of a, possibly infinite, series of sine's and cosines. This total is additionally taken as a Fourier expansion. The important disadvantage of a Fourier expansion is that it's fully frequency resolution and no time resolution and it's advanced and has poor energy compaction (Is the ability to pack the energy of the spatial sequence into as few frequency coefficient's as possible, this is very important for image compression, we represent the signal in the frequency domain if compaction is high we only have to transmit a few coefficients instead of the whole set of pixels). This suggests that though we would be ready to verify all the frequencies in a signal, we are not aware when they are present. To beat this downside within the past decade many solutions are developed that are additional or less ready to represent a sign within the time and frequency domain at an equivalent time. The thought behind these time-frequency joint representations is to cut the signal of interest into many components then analyze the components singly. It's clear that analyzing a signal this fashion can provide additional data concerning the once and wherever of various frequency elements, however it results in a basic downside as well. Suppose that we wish to understand precisely all the frequency elements given at an explicit moment in time. We have a tendency to cut out fully this short time window a employing a Kronecker delta function, remodel it to the frequency domain and. The matter here is that cutting the signal corresponds to a convolution between the signal and therefore the cutting window. Since convolution within the time domain is same for multiplication within the frequency domain and since the Fourier transform of a Kronecker delta contains all doable frequencies the frequency elements of the signal are dirty out everywhere the frequency axis. If truth be told this case is that the opposite of the quality Fourier Transform since we have a tendency to currently have time resolution however no frequency resolution.



Discrete cosine transform (DCT) has become the most popular technique for image compression over the past many years. One of the major reasons for its popularity is its selection as the standard for JPEG. Discrete Cosine Transform's are most commonly used for non-analytical applications such as image processing and signal-processing, applications such as video conferencing, fax systems, video's, and High Definition TV. Discrete Cosine Transform's can be used practically on a matrix of any dimension. Mapping an image space into a frequency space is the most common use of DCTs. For example, video is usually processed for compression/decompression as 8 x 8 blocks of pixels. Large and small features in a video picture are represented by low and high frequencies. An advantage of the DCT process is that image features do not normally change quickly, so many DCT coefficients are either zero or very small and require less data during compression algorithms. DCTs are fast, easy and has large energy compaction (Energy compaction means that most of the important information of the image is stored or compacted in top left corner of the image) [15]. Wavelets are considered better than DCT when it comes to getting better results in compression. With DCT, the picture blocks can lose their crisp edges, whereas, with wavelets the edges are very well explained.

The disadvantage of frequency domain (DCT) stego algorithms is that the hidden message length is very small. Also image quality decreases very fast, as concealed message size increases. Such techniques are comparatively easier to crack.

To further enhance our understanding of the result's of embedding, we study the behavior of discrete wavelet coefficients (DWT). Farid et al [11, 12] have shown that wavelet domain can capture image characteristics, like whether or not an image is a natural image or a computer generated one or may be scanned one. They have shown that the feature vector given by them are often be used for universal steganalysis. Their aim was fully to find whether or not an image contains any kind of hidden information.

The wavelet transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the Fourier transform C. Valens [14]. In wavelet analysis the use of a fully scalable modulated window solves the signal-cutting drawback. The window is shifted on the signal and for each position the spectrum is calculated. Then this method is repeated again and again with a little shorter (or longer) window for each new cycle. At the end the result will be a group of time-frequency representations of the signal, all with totally different resolutions. Due to this collection of representations we will speak of a multi resolution analysis. Within the case of wavelets we generally do not speak about time-frequency representations however about time-scale representations, scale being a method the alternative of frequency, as a result the term frequency is reserved for the Fourier transform.

### III. EMBEDDING AND EXTRACTION ALGORITHM

**3.1 Embedding Algorithm:** We are studying embedding algorithm & Extraction algorithm

Let the cover image be I, S is that the message bits to be hide. Initially we calculate the minimum value so that all the message bits are often embedded. Then, message digits are linearly hided into pairs of pixels. The detailed procedure is listed as below.

Input: Cover Image I, Data D, Key K.

Output: Stego Image.

Step1. Get the Image and convert it into byte stream.

Step2. Generate RGB and Hue Saturation Value (HSV) map.

Step3. Generate the Pixel map.

Step4. Call the optimize function.

Step5. Map co-ordinate with x, y.

Step6. Generate the sequences.

Step7. Calculate modulus distance and replace with other respective pixel.

Step8. Repeat step7 till completed.

Step9. Forward for LWZ out generate function.

**3.2 Extraction Algorithm:** To obtain the hidden message digits, pixel pairs are scanned within the same order as in the embedding procedure. The hidden message digits are the values of extraction function of the pixel pairs. The brief detailed is listed as below:

Input: Stego Image.

Output: Secret bit Stream.

Step1. Get input Image.

Step2. Generate byte stream.

Step3. Generate pixel to coordinate map.

Step4. Construct the sequences.

Step5. Provide input keys.

- Step6. Select mapped pixels.  
Step7. Calculate the distance.  
Step8. Repeat above step till output is obtained.  
Step9. Combine input stream and generate for LWZ analysis.

#### IV. CONCLUSION

The studied stegano graphic method known as LSB matching allows an embedding of the same amount of information into the stego image. At the same time, the number of changed pixel values is smaller. When more number of information is hidden, the appearance of image degrades. There are some drawbacks of Fourier transform such as it is complex and has poor energy compaction and Cosine also has some drawbacks as hidden message length is small and image quality degrades very fast, as embedded message size increases. Such techniques are comparatively easier to crack.

Thus all the above drawbacks are overcome by Wavelet Transform. Hence Wavelet Transform is the mostly used Transform and it also requires less time.

#### REFERENCES

- [1] "A Novel Data Embedding Method Using Adaptive Pixel Pair Matching", Wien Hong and Tung-Shou Chen
- [2] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [3] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography," *IEEE Security Privacy*, vol. 3, no. 3, pp. 32–44, May/Jun. 2003.
- [4] A. Cheddad, J. Condell, K. Curran, and P. McKeivitt, "Digital image steganography: Survey and analysis of current methods," *Signal Process.*, vol. 90, pp. 727–752, 2010.
- [5] T. Filler, J. Judas, and J. Fridrich, "Minimizing embedding impact in steganography using trellis-coded quantization," in *Proc. SPIE, Media Forensics and Security*, 2010, vol. 7541, DOI: 10.1117/12.838002.
- [6] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 1, pp. 111–119, Mar. 2006.
- [7] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," in *Proc. Int. Workshop on Multimedia and Security*, 2001, pp. 27–30.
- [8] A. D. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Process. Lett.*, vol. 12, no. 6, pp. 441–444, Jun. 2005.
- [9] C. K. Chan and L. M. Cheng, "Hiding data in images by simple LSB substitution," *Pattern Recognit.*, vol. 37, no. 3, pp. 469–474, 2004.
- [10] J. Wang, Y. Sun, H. Xu, K. Chan et.al, H. J. Kim, and S. H. Joo, "An improved section-wise exploiting modification direction method," *Signal Process.*, vol. 90, no. 11, pp. 2954–2964, 2004.
- [11] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [12] S. Lyu and H. Farid, "Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines," in 5th international workshop on Information Hiding, 2002.
- [13] Farid H, "Detecting Steganographic Message in Digital Images," Report TR2001-412, Dartmouth College, Hanover, NH, 2001.
- [14] C. Valens, "A Really Friendly Guide to Wavelets," available in <http://pagesperso-orange.fr/polyvalens/clemens/wavelets/wavelets.html>.
- [15] Zixiang Xiong, Kannan Ramchandran, Michael T. Orchard, and Ya-Qin Zhang; "A comparative study of DCT and wavelet based image coding," *IEEE Transactions on circuits and systems for video technology*, VOL. 9, NO. 5, AUGUST 1999.